

# A STUDY IN THE ANALYSIS OF STATIONARY TIME SERIES

BY

HERMAN WOLD

UPPSALA 1938  
ALMQVIST & WIKSELLS BOKTRYCKERI-A.-B.



## PREFACE.

In a sequence of fundamental memoirs, G. UDNY YULE, the eminent English statistician, has proposed certain methods of time series analysis which are of an essentially wider scope than the classical methods used in the search for periodicities. The basis of the new methods is a concept of flexible periodicity which in an ideal case reduces to the classical, functionally rigid periodicity. The importance and the broad applicability of the new ideas has been stressed particularly in subsequent discussion of the nature of business cycles.

In the recent rapid development of the theory of probability, the production of A. KHINTCHINE and A. KOLMOGOROFF represents a genuine discontinuity. A firm, axiomatic foundation has been obtained for the theory; other important results belong to the theory of random processes, i. e. hypothetical models for the analysis of time series. In accordance with the great diversity of time series, the main types of random process are of quite different structure.

In the theory of probability, the approaches of G. U. YULE fall under the heading of the stationary random process as defined and studied by A. KHINTCHINE. The present work might be described as a trial to subject the fertile methods of empirical analysis proposed by YULE to an examination and a development by the use of the mathematically strict tools supplied by the modern theory of probability. This statement, however, implies no valuation of the results and should be regarded rather as a tribute to my sources of inspiration and to the traditions of my milieu of study.

My most sincere thanks are due to my teacher, Professor HARALD CRAMÉR. His brilliant courses, distinguished by a spirit of realism combined with penetrating logic, have laid the foundation for my further work. As far as the present thesis is concerned, this is true not only in general but also in respect to particular parts

thereof, as indicated by the references to his 1933 course on Time Series Analysis. I wish to evidence my deep gratitude to Professor CRAMÉR also for the encouragement and interest shown me at all times, and culminating in his detailed reading of the first version of the manuscript. Our subsequent discussions have caused a revision particularly of the treatment of questions of convergence in probability.

To the Royal Swedish Academy of Sciences I want to express my respectful gratitude for a generous grant covering a substantial part of the expenses for printing and numerical calculation.

I am greatly indebted to my friends and colleagues Mr. G. ELFVING, Mr. W. FELLER and Mr. O. LUNDBERG for numerous stimulating discussions and for having read the manuscript and corrected many errors. I have also profited to a great extent by consultations with a large number of research workers in the different fields touched upon in the thesis. These obligations are, however, too comprehensive and indefinite to be expressed in detail.

Miss MARIA GISELSSON has assisted in making the model series and in the time-absorbing experiments for finding out reliable and convenient methods of applied analysis. Mr. L. JURÉN has performed the computations involved in applying the ultimate methods as presented in Chapter IV. Mr. B. BRUNO has furnished the correlograms dealt with in the latter part of section 31. Mrs. RUTH LEE-SANDSTRÖM has revised the language and assisted in reading the proofs. To all of these I extend my thanks.

Stockholm, July 1938.

H. W.

## Contents.

### Introduction.

	Page
1. Remarks on the scope of the study . . . . .	1
2. Principles of notation . . . . .	8

### CHAPTER I. A survey of hypotheses and methods proposed for the analysis of time series.

3. Scope and disposition . . . . .	13
4. Functional schemes . . . . .	14
5. On applied harmonic analysis . . . . .	17
6. On the linear difference equation . . . . .	19
7. A purely probabilistic approach . . . . .	21
8. A scheme of hidden periodicities . . . . .	22
9. On the criticism of the scheme of hidden periodicities . . . . .	25
10. Remarks on the schemes of linear regression . . . . .	27

### CHAPTER II. On the theory of the discrete stationary random process.

11. Definition of the stationary processes . . . . .	31
12. A theorem of A. KHINTCHINE . . . . .	35
13. Some fundamental operations with random processes . . . . .	35
14. On singular stationary processes . . . . .	41
15. Some type cases of the discrete stationary process . . . . .	47
16. On the normal stationary process . . . . .	60
17. The autocorrelation coefficients as FOURIER constants . . . . .	66
18. On linear approximation in a space of random variables . . . . .	75
19. Linear autoregression analysis of the discrete stationary process . . . . .	80
20. A canonical form of the discrete stationary process . . . . .	84

### CHAPTER III. On the theory of some special stationary processes.

21. On the concept of stochastical difference equation . . . . .	93
22. Some fundamentals concerning stochastical difference equations . . . . .	96
23. On the stationary processes with finite dispersion and with no singular component . . . . .	99
24. On the process of linear autoregression. General developments . . . . .	103
25. On some special cases of linear autoregression . . . . .	110
26. On the process of moving averages . . . . .	121

	Page
CHAPTER IV. On the application of some stationary schemes.	
27. Preliminary remarks. Disposition . . . . .	133
28. On the CRAIG effect . . . . .	135
29. On earlier applications of the scheme of linear autoregression	140
30. Preliminary survey of methods . . . . .	146
31. Some applications of the scheme of moving averages . .	150
32. Some applications of the scheme of linear autoregression .	174
APPENDIX A. On the $\omega^2$ -method for testing goodness of fit .	191
APPENDIX B. On the quantitative significance of correlation coefficients . . . . .	194
List of References . . . . .	208
Index . . . . .	212

## Introduction.

### 1. Remarks on the scope of the study.

Observational series which describe phenomena changing with time may be roughly classified in two broad categories, viz. *evolutive* and *stationary*. In the former case, different sections of the time series are dissimilar in one or more respects. For instance, the sectional averages may be distinctly different, or some other structural property of the series may present variation. In the analysis of evolutive time series, time evidently must be allotted a fundamental role, e. g. as the independent variable in a trend function, or as an absolute scale in studying the development of a phenomenon from an initial state of rest.

Stationary time series are unchanging in respect to their general structure. The fluctuations up and down in such a series may seem random or show tendencies to regularity — in any case, the character of the series is, on the whole, the same in different sections. Even without preparation, observational time series are frequently stationary. On the other hand, the deviations from a trend form a type of derived time series which is often stationary.

In the analysis of stationary time series, time plays the secondary role of a passive medium. Judging from the literature on the subject, the analysis of stationary time series might seem equivalent to the search for periodicities. In the present volume, let this be said at once, time series analysis is taken in a much wider sense.

Considering the classical methods of FOURIER and SCHUSTER, the hypothesis underlying these methods is that the time series under analysis might contain hidden periodicities, that is functional components which are periodic in the strict mathematical sense. It is well-known that these methods have often been applied with

definite success, and equally well-known that in many fields of scientific research they have met a severe criticism. An essential point in the criticism is that the idea of strict periods cannot possibly be realistic and adequate in certain applications. It has been claimed, for example, that in the theory of business cycles the hypothetical approach must be flexible to some extent, admit small changes in the periods and the amplitudes etc. A modified type of approach thus being called for, the difficulty is to find a precisely defined combination of the ideas of periodicity and of flexibility. It is evident that a strict hypothetical set-up as required can be reached only on the basis of the theory of probability.

Though the above mentioned critical argument is old, it was not until rather recently that approaches have been suggested which allow for changes in the waves in the time series under analysis. There are two main lines of approach, both of them germinating from G. U. YULE. Let these be briefly outlined.

Starting from a purely random series as given, for example, by dice-throwing, G. U. YULE (1921), (1926) forms the differences of a fixed order, and finds that the series thus obtained presents a tendency to regular fluctuations. E. SLUTSKY ((1927), (1937)) studies the effect of more general linear operations, and finds that under certain circumstances the resulting series will present sinusoidal waves with slowly changing amplitude and phase, waves showing a puzzling likeness to the cycles in economic time series. Nice examples of this are given by suitable moving averages of the primary random series. In the terminology of this study, the approaches mentioned are special cases of the scheme of *moving averages*.

The second type of approach is introduced by G. U. YULE (1927) in a study on sunspot numbers. Considering the sunspot index in a set of equidistant time points, YULE investigates the multiple correlation between these observations, and approximates by the use of linear regression analysis each observation by a linear function of the next preceding ones. The scheme thus implicitly defined will be called the scheme of *linear autoregression*. Using a physical interpretation, YULE gives a suggestive illustration of the new idea — a pendulum subjected to a stream of random shocks will be ruled by a scheme of linear autoregression. To a certain extent, the movement of the pendulum will bear resemblance to a free swinging, but the random impulses will cause a continuous shift in amplitude and phase.

Using a comprehensive term, the two schemes mentioned will be called schemes of *linear regression*. It is seen to be a common feature of these schemes that a random element plays a fundamental, active role. This constitutes a distinct contrast to the scheme of *hidden periodicities* — as we shall call the hypothesis of strict periods — and makes the schemes of linear regression *a priori* plausible in several instances where the scheme of hidden periodicities has been criticized.

From the viewpoint of the theory of probability, the schemes of linear regression are special cases of the *stationary random process* as defined and studied by A. KHINTCHINE ((1932)—(1934)). Let us discuss the situation in some detail.

Considering a phenomenon as described by an observational time series, let us fix arbitrarily a finite set of time points, say  $(t) = (t_1, t_2, \dots, t_n)$ . In a probabilistic theory of the phenomenon, we must necessarily assume that the behaviour of the time series in the  $n$  points considered is ruled by a definite probability distribution in  $n$  dimensions. Generally speaking, this distribution may be taken to be defined by a distribution function, say  $F(t_1, \dots, t_n; u_1, \dots, u_n)$ , where  $u_1, \dots, u_n$  are real variables. For instance, considering a set consisting of but one time point  $(t_1)$ , the hypothetical function  $F(t_1; u_1)$  will indicate the probability that the observational value in  $t_1$  is less than or equal to  $u_1$ . Having stated this, it is clear that we must assume certain relations of consistency between the functions  $F$  which belong to different sets  $(t)$ ; otherwise the hypothetical set-up might contradict itself. For instance, it is evident that  $F(t_1; u_1)$  must be assumed to satisfy all relations of the type  $F(t_1; u_1) = F(t_1, t_2; u_1, \infty)$ .

We have seen that the probabilistic treatment of a time series requires a set of distribution functions, say  $\{F\}$ , such that there is one function  $F$  corresponding to every finite set  $(t)$  of time points, and that the functions  $F$  satisfy certain consistency relations. On the other hand, such a hypothetical set-up will give a sufficient basis for a formal probabilistic analysis. Any set  $\{F\}$  with properties as mentioned is called a *random process*, and according to a fundamental theorem of A. KOLMOGOROFF (1933) such a set  $\{F\}$  is equivalent to a probability distribution in an infinite number of dimensions. Of course, each time point corresponds to one dimension in this distribution.

In defining a random process, we may either choose our points

( $t$ ) quite freely or else restrict them somehow, e. g. to an unbroken sequence of equidistant time points. Again following A. KOLMOGOROFF, a random process as defined by the set  $\{F\}$  is in the first case called *continuous*, in the case of equidistance *discrete*. Interpreting the random processes as random variables in an infinite number of dimensions, it is seen that the number of dimensions is enumerative in the case of a discrete process, and non-enumerative in the discrete case.

In applying a probabilistic scheme to a statistical distribution — perhaps multidimensional — each individual observation is looked upon as a sample value belonging to the hypothetical distribution in one or more dimensions. In the same manner, in interpreting an observational time series as belonging to a random process, the series is regarded as a sample value of the corresponding distribution in an infinite number of dimensions. Since a whole time series thus constitutes but one element in the statistical population, it is evident that the unconditioned random process is far too general to be useful in practical applications. According to the structure of the time series under analysis, we have to apply special types of the random process.

A. KHINTCHINE's definition of the stationary random process runs as follows. Letting  $(t) = (t_1, t_2, \dots, t_n)$  represent an arbitrary set of time points, and fixing arbitrarily a translation in time of this set, say  $(t^*) = (t_1 + t, t_2 + t, \dots, t_n + t)$ , a random process as defined by a set  $\{F\}$  of distribution functions is called stationary if the two functions  $F$  belonging to the two sets  $(t)$  and  $(t^*)$  are identical. Thus, the probability laws assumed to rule the observational time series depend on time in such a way that if we replace time as measured from a fixed time point by a time variable measured from another time point, the probability laws will remain the same. In other words, if the development in a time series is known up to a certain time point, say  $t$ , the probability laws ruling the continued development will depend only on the behaviour of the time series up to the time point  $t$ , not on the actual value of  $t$ . It can be seen that the postulated relativity of time is precisely adequate in view of the broad class of stationary time series roughly characterized above.

The present study is exclusively concerned with the theory and the applications of the *discrete stationary processes*. Accordingly, problems concerning evolutive time series fall outside the program. For instance, trend analysis will not be dealt with at all. Further,

evolutive random processes, and continuous stationary processes, will be touched upon only incidentally.

As far as I know, A. KHINTCHINE is alone in having dealt with the discrete stationary process in full generality; his chief result is that the stationary processes are ruled by a law of great numbers. The present study being concerned with other aspects of the stationary process, we shall next give a few comments on the main lines followed.

Chapter I serves a double purpose. In surveying the leading methods in the search for periodicities, particular attention is drawn to the hypotheses underlying these methods. By thus pointing out the rather narrow basis of the methods considered, the need for other types of hypothetical scheme is made clear, and the analysis of other types of approach prepared. On the other hand, after this rather detailed survey, the hypothesis of hidden periodicities will need no further separate treatment.

Chapter II is reserved for a general analysis of the discrete stationary process. Sections 13 and 14 are preparatory, and show that in certain respects the stationary processes may be dealt with in the same manner as random variables in a finite number of dimensions. In particular, the singular case considered in section 14 corresponds to a multi-dimensional probability distribution which is entirely concentrated in a plane or some other linear sub-space.

The discrete stationary process being extremely general, sections 15 and 16 give a few examples of processes obtained by different specializations. In this way we arrive at strict definitions of the processes of linear regression which cover the above mentioned approaches suggested by G. U. YULE. Further, the scheme of hidden periodicities is obtained by means of a singular stationary process. Detailed illustrations of the processes thus defined are given through model time series, i. e. series constructed in an artificial way on the basis of random sampling numbers. Finally, in section 16, the normal stationary process is defined by a straightforward generalization of the normal distribution in a finite number of dimensions.

In the latter part of the second chapter, the structural properties of the general stationary process are studied. The field being wide and unexplored, the analysis has been restricted to the elementary features. Generally speaking, only such properties have been taken into consideration which could be studied by the use of linear

operations and autocorrelation coefficients. As the autocorrelation coefficients correspond to the mixed moments of second order in a set of ordinary, one-dimensional variables, the analysis will, in certain respects, be parallel to the familiar theory of multiple correlation.

In section 17 is shown that the autocorrelation coefficients of a discrete stationary process may always be interpreted as the FOURIER coefficients of a non-decreasing function. This theorem discloses clearly that it is only in special cases that a periodogram can tell us something relevant about a time series.

The linear autoregression analysis as prepared in section 18 and developed in section 19 is based on the idea of subjecting the discrete stationary process to a treatment which is parallel to a time series analysis by means of the methods proposed by G. U. YULE (1927) and already referred to. The autoregression analysis thus corresponds to a linear regression analysis in a finite set of one-dimensional variables. As compared with periodogram analysis, the autoregression analysis is more general and efficient. For instance, considering the forecasts delivered by the different methods, the autoregression analysis reaches the limit beyond which we cannot proceed when employing only linear methods.

Using the same tools as in section 19, the analysis of the structure of the discrete stationary process is, in section 20, carried further in a quite new direction. In spite of its wide comprehensiveness, the general discrete stationary process is found to be of a readily surveyable structure. In fact, the general process is built up by two components which may be interpreted as generalized processes of hidden periodicities and of linear regression respectively.

In point of principle, the methods used in Chapter II are of a scope which would admit of generalizations of the analysis in different directions. Using non-linear operations, it would be possible to perform an autoregression analysis corresponding to curvilinear regression analysis in the case of a finite number of variables. Further, the analysis might be extended to the case when the time series is multi-dimensional, i. e. when several properties of the phenomenon observed are studied simultaneously.

The stochastical difference equations studied in the first two sections of Chapter III form a generalization of the ordinary linear difference equations. While the solutions of the latter equations are ordinary functions, the solutions of the former equations are

discrete random processes. Considering an ordinary linear difference equation, its solution describes how a certain oscillatory mechanism will develop under given conditions. On the other hand, if a primary stream of impulses is defined by means of a random process, the solutions of a corresponding stochastical difference equation give the probability laws which will rule the oscillatory mechanism when subjected to the primary impulses. Of course, if the actual development of the mechanism is known, the corresponding series of impulses is readily obtained. Moreover, among the solutions of stochastical difference equations we find both evolutive and stationary random processes, e. g. the process of linear autoregression as strictly defined in section 15.

The latter part of Chapter III is reserved for a detailed study of the processes of linear regression, a chief purpose being to illustrate the general analysis in Chapter II, and the theory of stochastical difference equations. Particular attention is paid to the forecast situation. In contradistinction to the processes of hidden periodicities, the processes of linear regression contain an active random element which affects the efficiency of the forecast. As a matter of fact, in a process of linear regression, the efficiency is the less, the longer the interval of time forecasted over. On the other hand, the short time forecasts are the more efficient. In view of the applications this circumstance is advantageous, for, of course, the main interest is always focussed upon the short time forecasts.

The analysis in Chapter III, too, suggests certain generalizations. Thus, nothing prevents us from defining random processes by means of non-linear stochastical difference equations. Moreover, multi-dimensional random processes may be defined on the basis of systems of stochastical difference relations. A few remarks along the latter line are given in sections 31 and 32.

Chapter IV, finally, gives a few applications of the theoretical analysis to observational time series of the stationary type. Such a series being given, its correlogram — my term for the auto-correlation periodogram — is adopted as an indicator of which type of process to apply. Explicit applications being given of the processes of linear autoregression and of moving averages, general methods are indicated for finding suitable numerical values of the parameters involved. For instance, assuming a given time series to be a moving average of an unknown primary series, suitable values for the weights of the hypothetical moving average are derived

from the correlogram of the given series. Further, a general method is given for deriving the primary series which corresponds to such a set of weights.

The purpose of the applications is to illustrate certain general methods of analysis, not to supply a theory of the phenomena described by the time series examined. Accordingly, I make little account of the hypothetical schemes arrived at, and no attempts to test the significance of the parameters determined. On the contrary, warnings are repeatedly given for attaching importance to the numerical results of the analysis, for one reason because significance questions are extremely intricate in time series analysis. Another reason for caution is discussed in Appendix B, where it is pointed out that the empirical correlation coefficients used in time series analysis are quantitatively conditioned by the size of the statistical mass under observation.

## 2. Principles of notation.

The present volume being concerned with both theory and applications, symbols are needed for probabilistic as well as statistical concepts. Since the purpose of this section is to indicate the principles of notation, no completeness is aimed at with respect to the definitions of the elementary concepts considered. For these, reference is made to G. U. YULE and M. G. KENDALL (1937) as to the theory of statistics, and to H. CRAMÉR (1937) as to the theory of probabilities.

Generally speaking, the notations try to bring into relief the correspondence between theoretical and empirical concepts. Thus, for parallel theoretical and empirical concepts the same symbol will be used — in the latter case marked by a bar. As a rule, Greek letters will be reserved for random variables, Roman letters for functions, ordinary variables, and constants.

In agreement with these general principles, random variables as dealt with in statistics will be denoted by  $\xi$ ,  $\bar{\eta}$ , etc. Considering such a random or statistical variable, say  $\xi$ , its observational values in a particular statistical population will be distinguished by running indices, e. g.

$$(1) \quad \xi_1, \bar{\xi}_2, \dots, \bar{\xi}_n.$$

Irrespective of the order of the elements, an observational set (1) is uniquely determined by the corresponding function of cumulative relative frequencies, say  $\bar{F}(u)$ . Thus  $\bar{F}(u)$ , for every real  $u$ , equals the relative number of statistical units with an observational value  $\xi_i \leq u$ . The function  $\bar{F}(u)$  will be called the (empirical) distribution function of the observational set. By the use of the STIELTJES integral, the elementary characteristics of an empirical distribution (1) can be conveniently expressed in terms of  $\bar{F}(u)$ . The average of  $\bar{\xi}$  in the population considered is

$$\bar{m} = \frac{1}{n} \cdot \sum_{i=1}^n \bar{\xi}_i = \int_{-\infty}^{\infty} u \cdot d\bar{F}(u).$$

The central moment of order  $k$  is denoted by  $\bar{\mu}_k$ , and reads

$$\bar{\mu}_k = \frac{1}{n} \cdot \sum_{i=1}^n (\bar{\xi}_i - \bar{m})^k = \int_{-\infty}^{\infty} (u - \bar{m})^k \cdot d\bar{F}(u).$$

Thus,  $\bar{\mu}_2$  is the variance. Denoting the dispersion by  $\bar{D}$ , we have

$$(2) \quad \bar{D}^2 = \bar{\mu}_2.$$

In studying a statistical variable  $\bar{\xi}$ , we introduce a corresponding hypothetical random variable  $\xi$ . Such a variable  $\xi$ , which is also called stochastical or aleatory, is completely characterized by its distribution function  $F(u)$ . By definition,  $F(u)$  indicates the probability that  $\xi$  is less than or equal to  $u$ . This is expressed by the relation

$$(3) \quad F(u) = P[\xi \leq u],$$

where  $P$  is the probability function of  $\xi$  (cf. fig. 18).

In the analysis of a variable  $\bar{\xi}$ , it is actually only the observational values  $\bar{\xi}_i$ , and the variable  $\xi$  as defined by the distribution function  $F(u)$ , which will appear in the formal developments. However, in the literature the observational values  $\bar{\xi}_i$  are also often supplied with hypothetical parallels, called sample values. For instance, the distribution of  $\xi$  is spoken of as constituted by an infinite population of sample values  $\xi_i$ . Such a terminology is often convenient, and is, incidentally, used also in the present study. As we need no notation for these sample values, the symbol  $\xi_i$  will,

as indicated later on, be reserved for representing independent variables in random sampling.

Let  $g(x)$  be a function of a real variable, and  $\xi$  a random variable with distribution function  $F(u)$ . Under general conditions,  $g(\xi)$  represents a random variable with expectation given by

$$(4) \quad E[g(\xi)] = \int_{-\infty}^{\infty} g(u) \cdot dF(u).$$

In particular, the elementary characteristics of  $\xi$  may be interpreted in this way. For instance, the mean ( $m$ ), dispersion ( $D$ ), and the central moments ( $\mu$ ) are given by

$$(5) \quad m = E[\xi] = \int_{-\infty}^{\infty} u \cdot dF(u), \quad D^2(\xi) = \mu_2.$$

$$(6) \quad \mu_k = E[(\xi - m)^k] = \int_{-\infty}^{\infty} (u - m)^k \cdot dF(u).$$

Any multi-dimensional random variable may be looked upon as a combination of one-dimensional variables. For such variables we shall use notations of type  $\xi = [\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(n)}]$ . Interpreting in this manner, we can often use the same notations as in the one-dimensional case. When full information is required, e. g. concerning (3), we shall use notations such as

$$F(u_1, u_2, \dots, u_n) = P[\xi^{(1)} \leq u_1, \xi^{(2)} \leq u_2, \dots, \xi^{(n)} \leq u_n].$$

In the same way, the expression (4) may be regarded as the expectation of a function  $g(\xi)$  of a multi-dimensional variable  $\xi$ , only that  $E[g(\xi)]$  must be interpreted as a vector in the space of  $g(\xi)$ , and that the integral must be extended over the space of  $\xi$ . For instance, let a one-dimensional random variable  $p(\xi)$  be defined as a function of an  $n$ -dimensional variable  $\xi$  with distribution function  $F(u) = F(u_1, \dots, u_n)$ . Then the distribution function of  $p(\xi)$ , say  $G_p(x)$ , will equal, for an arbitrarily fixed  $x$ , the expectation of a function  $g_{x,p}(u) = g_{x,p}(u_1, \dots, u_n)$  defined by the relations  $g_{x,p}(u) = 1$  as  $p(u) \leq x$ , and  $g_{x,p}(u) = 0$  as  $p(u) > x$ . Thus,

$$G_p(x) = E[g_{x,p}(\xi)] = \int_{R_n} g_{x,p}(u_1, \dots, u_n) \cdot dF(u_1, \dots, u_n).$$

In particular, letting  $F(u_1, \dots, u_n)$  be the distribution function of

a variable  $\xi = [\xi^{(1)}, \dots, \xi^{(h)}]$ , and  $G_p(x)$  that of the sum  $p(\xi) = p_1 \xi^{(1)} + p_2 \xi^{(2)} + \dots + p_h \xi^{(h)}$ , where the  $p_i$ 's are real constants, we have

$$(7) \quad G_p(x) = \int_{p(x)} d_{u_1, \dots, u_h} F(u_1, \dots, u_h),$$

the integration being extended over a half-space  $p(x)$  defined by  $p(x) = [p_1 u_1 + \dots + p_h u_h \leq x]$ . The present study being chiefly concerned with linear functions of random variables, this formula will frequently come into use.

Considering a combined variable  $\xi = [\xi^{(1)}, \dots, \xi^{(h)}]$  with distribution function  $F(u_1, \dots, u_h)$ , and taking  $p_k = 0$  for  $k \neq i$ , and  $p_i = 1$ , formula (7) gives the distribution function of the individual variable  $\xi^{(i)}$ . Denoting the resulting functions by  $F_i(u)$ , the variables  $\xi^{(i)}$  are called *independent* if, for an arbitrary set  $(u_1, \dots, u_h)$ , the following relation is satisfied,

$$F(u_1, u_2, \dots, u_h) = F_1(u_1) \cdot F_2(u_2) \dots F_h(u_h).$$

In the case of independent variables, the expectation satisfies a general relation involving arbitrary functions  $g_i$ , viz.

$$(8) \quad E[g_1(\xi^{(1)}) \cdot g_2(\xi^{(2)}) \dots g_h(\xi^{(h)})] = E[g_1(\xi^{(1)})] \cdot E[g_2(\xi^{(2)})] \dots E[g_h(\xi^{(h)})].$$

Putting  $p_i = 1$  for all  $i$ , formula (7) gives the distribution function, say  $G(u)$ , of the sum of our  $h$  variables  $\xi^{(i)}$ . If these are independent,  $G(u)$  can be expressed by the use of the familiar *composition (convolution, Faltung) symbol  $*$* , viz.

$$(9) \quad G(u) = F_1(u) * F_2(u) * \dots * F_h(u).$$

In the case of two independent variables, the convolution is given by

$$G(u) = F_1(u) * F_2(u) = \int_{-\infty}^{\infty} F_1(u - x) \cdot d F_2(x).$$

A *random sample* containing  $h$  elements, and belonging to a random variable  $\xi$  with distribution function  $F(u)$ , may be defined as a sample value of an  $h$ -dimensional variable obtained by combining  $h$  independent variables  $\xi_i$  which have the same distribution function  $F(u)$ . The concept of a random sample is thus purely theoretical, and corresponds to that of a statistical population. As already men-

tioned, we need no particular notation for the elements in a random sample.

The ordinary correlation coefficient between two interdependent random variables, say  $\xi(i)$  and  $\xi(k)$ , will be denoted  $r = r(\xi(i); \xi(k))$ .

Let  $\dots, \xi(-1), \xi(0), \xi(1), \xi(2), \dots$  be a sequence of interdependent random variables such that the characteristics

$$(10) \quad m(\xi; t) = E[\xi(t)]; \quad \nu_2^{(k)}(\xi; t) = E[\xi(t) \cdot \xi(t+k)]$$

will exist for all integral  $t$  and  $k$ . Let it further be assumed that the quantities defined by

$$(11) \quad \begin{cases} m(\xi) = \lim_{\substack{n \rightarrow \infty \\ n' \rightarrow -\infty}} \frac{1}{n - n' + 1} \cdot \sum_{t=n'}^n m(\xi; t); \\ \nu_2^{(k)}(\xi) = \lim_{\substack{n \rightarrow \infty \\ n' \rightarrow -\infty}} \frac{1}{n - n' + 1} \cdot \sum_{t=n'}^n \nu_2^{(k)}(\xi; t) \end{cases}$$

will exist. Under these circumstances, the coefficients  $r_k$  defined by

$$(12) \quad r_k = r_k(\xi) = \frac{\nu_2^{(k)} - m^2}{\nu_2^{(0)} - m^2}$$

will be called the *autocorrelation coefficients* of the sequence  $\xi(t)$ . It should be observed that this definition holds also in the special case when  $\xi(t)$  reduces to an ordinary function of  $t$ .

Let (1) represent an empirical time series obtained by observations in the equidistant time points  $t = 1, 2, \dots, n$ . Following G. U. YULE (1926), the coefficients  $\bar{r}_k$  defined by

$$(13) \quad \bar{r}_k = \frac{\sum_{i=1}^{n-k} \bar{\xi}_i \cdot \bar{\xi}_{i+k} - (n-k) \cdot \bar{m}_1 \cdot \bar{m}_2}{(n-k) \cdot \bar{D}_1 \cdot \bar{D}_2},$$

where

$$\bar{m}_1 = \frac{1}{n-k} \cdot \sum_1^{n-k} \bar{\xi}_i; \quad \bar{m}_2 = \frac{1}{n-k} \cdot \sum_{k+1}^n \bar{\xi}_i,$$

and

$$\bar{D}_1^2 = \frac{1}{n-k} \cdot \sum_1^{n-k} (\bar{\xi}_i - \bar{m}_1)^2; \quad \bar{D}_2^2 = \frac{1}{n-k} \cdot \sum_{k+1}^n (\bar{\xi}_i - \bar{m}_2)^2,$$

will be called *serial coefficients*. These obviously form an empirical parallel to the autocorrelation coefficients.

## CHAPTER I.

### A survey of hypotheses and methods proposed for the analysis of time series.

#### 3. Scope and disposition.

This chapter aims at giving a historical perspective to the investigations in the subsequent chapters. Within the bounds of the survey fall the fundamental facts concerning the principal theoretical schemes set up for the study of stationary time series in one dimension. The leading methods for fitting the considered schemes to observational data will also be examined. For the sake of concreteness some descriptive methods will be touched upon incidentally. The survey being concerned with a general outline only, reference is given to H. BURKHARDT (1904) and K. STUMPF (1927), (1937) for completion.

The purely functional schemes will be dealt with first among the theoretical models for a given time series. At the opposite extreme is taken that purely probabilistic scheme in which the series is regarded as a random sample of an aleatory variable. The other schemes may be looked upon as intermediate cases. Of the mixed schemes, the approach of hidden periodicities is treated first. The series is here assumed to be additively composed of independent functional and random elements. The last section of the survey deals with the scheme of moving averages, studied in certain special cases by G. U. YULE (1921) and E. SLUTSKY (1927), and with the scheme proposed by G. U. YULE (1927) under the name of disturbed harmonics (cf. p. 2).

The time points considered will be equidistant. The unit of equidistance will be taken for the time unit, a simplification which evidently does not involve any loss of generality.

#### 4. Functional schemes.

The functional schemes aim at a perfect functional representation of the empirical data.

As a first example of functional approach, we take the hypothesis of a periodic function. Assuming the period to equal  $h$ , and denoting the hypothetical function by  $x(t)$ , the following relation will be satisfied for any  $t$

$$(14) \quad x(t) - x(t-h) = 0.$$

As  $t$  is given for integral values only, it will be sufficient to consider the case of an integral  $h$ . Then (14) forms an ordinary difference equation of order  $h$ . According to the theory of difference equations, the solutions of (14) may be written

$$(15) \quad x(t) = m + \sum_k C_k \cos \left( \frac{2\pi}{h} \cdot kt + \varphi_k \right) = \\ = m + \sum_k \left( A_k \cos \frac{2\pi}{h} kt + B_k \sin \frac{2\pi}{h} kt \right),$$

where  $k$  runs from 1 to  $(h-1)/2$  or  $h/2$ . The real parameters  $A_k, B_k, C_k > 0$ , and  $\varphi_k$  in  $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$  are connected by the relations

$$(16) \quad C_k^2 = A_k^2 + B_k^2, \quad \varphi_k = \arctg (-B_k/A_k).$$

The approach now described will be termed the *scheme of periodic functions*. According to (15), the approach function may be considered to be composed of superposed harmonics, each having an amplitude  $C_k$ , a phase  $\varphi_k$ , and a frequency  $\lambda_k$  given by

$$\lambda_k = \frac{2\pi}{h} \cdot k.$$

The expression (15) may be taken as the basis for various generalized hypotheses. We start with the approach

$$(17) \quad x(t) = m + \sum_{k=1}^s C_k \cos (\lambda_k t + \varphi_k) = \\ = m + \sum_{k=1}^s (A_k \cos \lambda_k t + B_k \sin \lambda_k t),$$

which will be called the *scheme of superposed harmonics*. Alternately, the function (17) will be called a *composed harmonic*.

Denoting by  $p_k$  the periods of the individual harmonics in (15) or (17), we have

$$(18) \quad p_k = 2\pi/\lambda_k.$$

In the scheme (15) the periods of the individual harmonics representing  $x(t)$  are seen to be true fractions of  $h$ . In the generalized scheme (17) this restriction will not be laid down on the individual periods as given by (18). On the other hand, since  $t$  takes on integral values only, it follows that for any integer  $n$  the substitution of  $\lambda_k - n \cdot 2\pi$  for  $\lambda_k$  will have no effect in (17). Neither will  $x(t)$  be affected by a simultaneous substitution of  $-\lambda_k$  for  $\lambda_k$ , and  $-\varphi_k$  for  $\varphi_k$ . Thus it would involve no loss of generality to assume that  $0 < \lambda_k \leq \pi$ . This means that, in point of principle, the analysis must be restricted to periods not less than 2 time units. A study of shorter periods requires a reduction of the time unit chosen as a basis for the analysis. However, unless explicitly stated otherwise, we shall merely assume that  $0 < \lambda_k$ .

A function  $x(t)$  of type (17) belongs, as is well known, to the class of almost periodic functions in the sense of H. BOHR (see e. g. (1925)). The following property of an almost periodic function  $x(t)$  is recorded for later use: An  $\varepsilon > 0$  being arbitrarily given, there exists for every number  $t_0 > 0$  an integer  $T(\varepsilon, t_0) > t_0$  such that for every  $t$  (see H. BOHR (1925), p. 88)

$$(19) \quad |x(t + T) - x(t)| < \varepsilon.$$

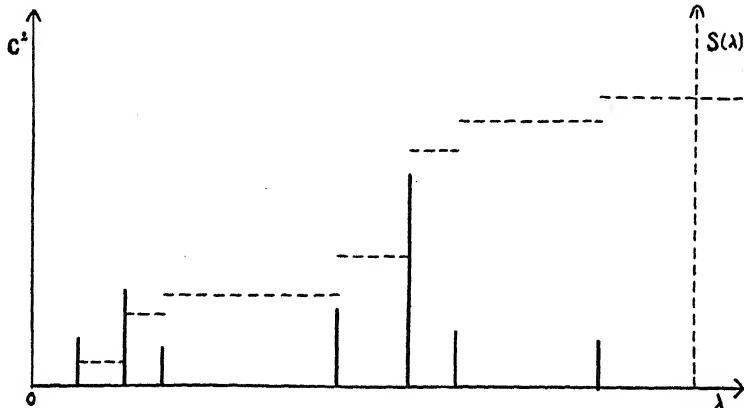
A rough description of a scheme of superposed harmonics is yielded by its periodogram. This has the frequency  $\lambda > 0$  for horizontal axis, and indicates by ordinates in  $\lambda_k$  the corresponding squared amplitudes  $C_k^2$ . It is seen that the periodogram — which is met in several variants, e.g. with the periods  $p_k = 2\pi/\lambda_k$  as abscissae or with the amplitudes  $C_k$  as ordinates — does not pay any regard to the constant term or to the phases appearing in the expression (17). Another variant is the integrated periodogram. This is a function, say  $S(\lambda)$ , defined by

$$(20) \quad S(\lambda) = \sum_{\lambda_k \leq \lambda} C_k^2 / \sum_{k=1}^s C_k^2.$$

Thus  $S(\lambda)$  is a step (or *saltus*) function which is proportionate to the

sum of squared amplitudes with frequency not greater than  $\lambda$ . An example of periodogram and corresponding integrated periodogram is given in the figure below.

In the study of light, the periodogram has an experimental parallel in the spectrum. The prism spreads the light waves according to their frequency  $\lambda_k$ , and the individual lines in the spectrum indicate the energy of respective wave components. This energy is proportionate to the squared amplitude. The energy represented in an interval of the spectrum is thus proportionate to the sum of the



*Fig. 1. Ordinary periodogram ( $C^2$ , vertical lines), and corresponding integrated periodogram ( $S(\lambda)$ , horizontal lines).*

ordinates in the corresponding interval of the ordinary periodogram, and proportionate to the increase of the integrated periodogram in the same interval.

Analysis of white light produced spectra, where the lines were thin and lying very close together. This fact gave rise to the idea of continuous spectra and periodograms, in so far as the energy belonging to an interval was thought of as the integral of a spectral density. A survey and a development of the mathematical theory used in this connexion has been given by N. WIENER (1930). Translating to the terminology of the present study, this theory is based upon an analysis of the function

$$(21) \quad Q(u) = \lim_{z \rightarrow \infty} \frac{1}{2z} \cdot \int_{-z}^z (x(t+u) - m) \cdot (x(t) - m) dt,$$

where

$$m = \lim_{z \rightarrow \infty} \frac{1}{2z} \int_{-z}^z x(t) dt.$$

It is seen that if  $x(t)$  is given by (17),  $Q(u)$  reduces to

$$(22) \quad \frac{1}{2} \cdot \sum_{k=1}^s C_k^2 \cdot \cos \lambda_k u.$$

Modifying by a constant factor the function  $S(\lambda)$  defined by WIENER, this reduces in the non-complex case to

$$(23) \quad S(\lambda) = \frac{2}{\pi} \cdot \int_0^\infty \frac{Q(u)}{Q(0)} \cdot \frac{\sin \lambda u}{u} du.$$

$S(\lambda)$  is called the *integrated periodogram* of  $x(t)$ , for in the case of superposed harmonics (17),  $S(\lambda)$  as given by (23) reduces to (20) (see e.g. H. C. CARSLAW (1930), p. 322). WIENER shows that  $S(\lambda)$  is always a non-decreasing function, and that there exist functions  $x(t)$  with continuous integrated periodograms (23).

N. WIENER applies the generalized harmonic analysis also to functions defined by a random scheme. For instance, if  $x(t)$  in integral intervals is 1 or  $-1$  with equal probability, and if the values taken on in different intervals are independent,  $S(\lambda)$  is with probability 1 given by

$$(24) \quad S(\lambda) = \frac{2}{\pi} \cdot \int_0^1 \frac{1 - \cos u}{u^2} du.$$

## 5. On applied harmonic analysis.

Let an observational time series be represented by (1). The problem is to choose a suitable hypothetical scheme, and to find for the parameters involved numerical values yielding as good fit as possible to the empirical data.

In case the observational data are strictly periodic, say with period  $p$ , an application of (15) by means of the FOURIER analysis will yield a perfect fit. It will be sufficient to consider the data  $\bar{\xi}_1, \dots, \bar{\xi}_p$  ranging over one period. However, since the formulae become particularly simple in case of an even period, and since we may take the double period for a basis if  $p$  is odd; let it be assumed that  $h = p = 2q$ . The FOURIER formulae for the  $2q$  parameters  $m, A_q, A_k, B_k$ , where  $k = 1, 2, \dots, q-1$ , then read as follows (see e.g. H. S. CARSLAW (1930), p. 325)

$$(25) \quad \begin{cases} m = \frac{1}{p} \cdot \sum_{t=1}^p \bar{\xi}_t, & A_q = \frac{1}{p} \cdot \sum_{t=1}^p \bar{\xi}_t \cos \pi t, \\ A_k = \frac{1}{q} \cdot \sum_{t=1}^p \bar{\xi}_t \cos \frac{\pi}{q} kt, & B_k = \frac{1}{q} \cdot \sum_{t=1}^p \bar{\xi}_t \sin \frac{\pi}{q} kt. \end{cases}$$

In the approach (17), the essential problem is to evaluate the frequency numbers  $\lambda_k$ . The principal method is that of A. SCHUSTER ((1898), (1900)) which is based upon the construction of an empirical

periodogram, say  $\bar{C}^2(\lambda)$ . The formulae required are (see e.g. K. STUMPF (1927), p. 103)

$$(26) \quad \begin{cases} A(\lambda) = \frac{2}{n} \cdot \sum_{t=1}^n (\bar{\xi}_t - \bar{m}) \cdot \cos \lambda t, & 0 < \lambda < \pi, \\ B(\lambda) = \frac{2}{n} \cdot \sum_{t=1}^n (\bar{\xi}_t - \bar{m}) \cdot \sin \lambda t, & 0 < \lambda < \pi; \end{cases}$$

$$(27) \quad \bar{C}^2(\lambda) = A^2(\lambda) + B^2(\lambda).$$

For  $\lambda = \pi$ , the factor 2 in  $A$  and  $B$  must be omitted.

A graph of the curve  $\bar{C}^2(\lambda)$  presents characteristic maxima, the abscissae  $\lambda_k$  of which are taken for the frequency numbers sought for. The corresponding parameters  $A_k = A(\lambda_k)$  and  $B_k = B(\lambda_k)$  are obtained from (26).

The periodogram method gives much valuable information about the series under investigation, but is rather inconvenient. However, a careful fit of (17) to observational data seems to necessitate tedious computations, so the labour seems due to the problem, not to the method. Nevertheless, many simplified and, accordingly, approximate methods have been proposed. One type of these is of interest for the sequel because it is based upon the differential or difference relations satisfied by a sum of harmonics. The first method of this kind, that of S. OPPENHEIM (1909), utilizes the fact that for arbitrary  $C_k$  and  $\varphi_k$  the function  $x(t)$  given by (17) satisfies a differential equation of  $2s$ :th order,

$$(28) \quad x^{(2s)}(t) + g_1 \cdot x^{(2s-2)}(t) + \cdots + g_{s-1} \cdot x^{(2)}(t) + g_s \cdot [x(t) - m] = 0,$$

with constant coefficients  $g_i$  such that the equation

$$(29) \quad z^s + g_1 \cdot z^{s-1} + \cdots + g_{s-1} \cdot z + g_s = 0$$

has the roots  $-\lambda_1^2, -\lambda_2^2, \dots, -\lambda_s^2$ . Identifying  $\bar{\xi}_t$  with  $x(t)$ , and taking  $m = \bar{m}$ , S. OPPENHEIM uses the identity (28) for  $s$  successive  $t$ -values, the  $g_i$  being so far undetermined. These relations are considered a system determining the  $g_i$ 's. Inserting the resulting  $g_i$ 's in (29), the solving of this equation gives, finally, the frequency numbers  $\lambda_k$  desired.

The derivatives required for the system of identities (28) S. OPPENHEIM obtains from the observational differences  $\Delta^{2i} \bar{\xi}_t$  by the

well-known serial development (see e.g. E. T. WHITTAKER and G. ROBINSON (1926), p. 64). The intricate passage from differences to differentials is avoided in the modification of the OPPENHEIM method given by H. BRUNS (1911), who starts from a certain identity between central differences satisfied by  $x(t)$ , viz.

$$(30) \quad \mathcal{A}^{2s}x(t-s) + h_1 \cdot \mathcal{A}^{2s-2}x(t-s+1) + \cdots + h_{s-1} \cdot \mathcal{A}^2x(t-1) + \\ + h_s[x(t) - m] = 0.$$

Here the constant coefficients  $h_i$  are such that the roots of the equation

$$z^s + h_1 \cdot z^{s-1} + \cdots + h_{s-1} \cdot z + h_s = 0$$

are  $-q_k^s$ , where

$$(31) \quad q_k = 2 \sin \lambda_k / 2.$$

The fact that different functional schemes may give a good fit gives rise to the question of which scheme should be preferred when analysing a given time series. This is a particular aspect of the general test problem which is fundamental in all practical applications. However, the most important aspects of the test problem belong to the theory of probability, and will be touched upon later. As purely functional test methods may be regarded those which in principle consist in an extension — extrapolation or, sometimes, interpolation — of the observational material, and a comparison with the corresponding values of the functions fitted to the original data.

## 6. On the linear difference equation.

A class of functions of importance in the sequel, though not as a scheme for time series, is formed by the solutions of linear difference equations with constant coefficients,

$$(32) \quad (x(t) - m) + a_1 \cdot (x(t-1) - m) + \cdots + a_h \cdot (x(t-h) - m) = 0, \quad a_h \neq 0.$$

Writing (30) on the form (32), the resulting sequence  $a_i$  will be symmetrical. Now, since (30) is a special case of (32), the solutions of (32) embrace (17), and are well-known to be (see e.g. P. M. MARPLES (1932))

$$(33) \quad x(t) = m + \sum_{k=1}^i H_{m_k-1}^{(1)}(t) \cdot p_k^t + \sum_{k=1}^j [H_{n_k-1}^{(2)}(t) \cdot \cos \lambda_k t + H_{n_k-1}^{(3)}(t) \cdot \sin \lambda_k t] \cdot q_k^t,$$

where  $H_r^{(s)}$  stands for a polynomial of order  $r$ . While the polynomials  $H$  are of arbitrary coefficients, their orders are the same in all solutions. In fact, the orders are determined by the *characteristic equation* of (32), viz.

$$(34) \quad z^h + a_1 \cdot z^{h-1} + \cdots + a_{h-1} \cdot z + a_h = \\ = \prod_{k=1}^i (z - p_k)^{m_k} \cdot \prod_{k=1}^j (z^2 + 2 s_k \cdot z + q_k^2)^{n_k} = 0,$$

where the factors in the second member are real.

The frequencies  $\lambda_k$  in (33) are connected with the characteristic equation by the relations

$$(35) \quad \cos \lambda_k = -s_k/q_k.$$

The asymptotical behaviour of  $x(t)$  is dependent on the exponential factors  $p_k^t$  and  $q_k^t$ , the bases of which are likewise seen to be uniquely determined by (34).

For later application it should be noticed that a necessary and sufficient condition for the convergence of  $\sum_{t=1}^{\infty} (x(t) - m)^2$  and  $\sum_{t=1}^{\infty} |x(t) - m|$ , for any values taken on by the arbitrary coefficients in the polynomials  $H$ , is that  $|p_k| < 1$  and  $|q_k| < 1$  for all  $k$ . Another wording of the condition is that all roots of the characteristic equation (34) shall lie within the periphery of the unit circle. In such a case,  $x(t)$  will be referred to as describing a *damped oscillation*.

A second property of (32) will also be used later. Let the arbitrary coefficients in (33) be fixed under the single condition that no polynomial  $H$  vanishes; then, if some  $p_k$  or some  $q_k$  is different from unity in modulus, formula (33) shows that  $|x(t)|$  cannot possibly be uniformly bounded in  $(-\infty < t < \infty)$ . In the same way, any solution which does not belong to an equation of lower order is unbounded if (34) presents a multiple root. Thus, if a solution

$x(t)$  of (32) satisfies no linear equation of lower order and if  $|x(t)|$  is uniformly bounded in  $(-\infty < t < \infty)$ , then  $|p_k| = |q_k| = m_k = n_k = 1$ , i. e. the equation (32) is of the special type (30), and  $x(t)$  is of type (17).

### 7. A purely probabilistic approach.

A typical probabilistic hypothesis about a given time series is to regard the observational values as a random sample of a certain aleatory variable. The complete hypothetical set-up thus consists of a sequence of random variables, say  $\dots, \eta(t-1), \eta(t), \eta(t+1), \dots$ , which are mutually independent, and have identical distribution functions, say  $F(x)$ .

Since the hypothesis under consideration consists of two elements, the test methods are of two kinds: a) those testing the goodness of fit of the hypothetical function  $F(x)$  to the empirical distribution function  $\bar{F}(x)$  characterizing the observational series, and b) those testing the randomness in the observational series.

A perfect fit to the data being in contrast to the idea of randomness, an amount of arbitrariness is in place in the choice of an hypothetical distribution function  $F(x)$  characterizing the aleatory variables  $\eta(t)$ .

The classical method for testing goodness of fit is the  $\chi^2$ -method of K. PEARSON (see e. g. G. U. YULE and M. G. KENDALL (1937), Chapter 22). Another method has been proposed by H. CRAMÉR ((1927) p. 112, and (1928) p. 145), and, under the term of  $\omega^2$ -method (Summenlinienverfahren), by R. v. MISES ((1930) p. 316). The latter method, an interesting modification of which has been given by N. SMIENOFF (1936), does not suffer from the well-known arbitrariness implied in the  $\chi^2$ -method. In Appendix A is given a labour-saving device, whereby the  $\omega^2$ -method is also made more convenient than the  $\chi^2$ -method.

The hypothetical randomness lies in the independence relations of the type (8). Accordingly, the tests of randomness are tests examining various particular instances of these relations. For example, since  $r_k = 0$  for  $k > 0$ , the serial coefficients  $\tilde{r}_k$  must approximate zero for  $k > 0$  (the particular case  $k = 1$  is the ABBE-HELMERT criterion — cf. K. STUMPF (1927), p. 8).

An important instance of the general test problem is concerned

with the choice between functional and probabilistic schemes. This problem is dealt with in the expectance theory (see e. g. K. STUMPF (1927), p. 115) founded by A. SCHUSTER (1898), (1900)) and which studies, i. a., the distribution of the periodogram ordinates  $C^2(\lambda)$  obtained when substituting a set of independent random variables  $\eta(t)$  (cf. p. 11) for (1) in (26). Taking  $\bar{m} = E[\eta]$  in (26), the basic formulae of the expectance theory read

$$(36) \quad E[A(\lambda)] = E[B(\lambda)] = 0, \quad 0 < \lambda \leq \pi.$$

$$(37) \quad E[C^2(\lambda)] = \frac{4}{n} D^2(\xi), \quad 0 < \lambda < \pi.$$

In case  $\lambda = \pi$ , the factor 4 in the latter formula must be omitted.

Having now touched upon some typical functional and probabilistic schemes set up for the analysis of time series, we are in a position to pass on to some intermediary schemes. Of these, two types may be distinguished which are of fundamentally different character. Since the terminology does not seem to be established, I propose for the two types the names »schemes of *hidden periodicities*» and »schemes of *linear regression*». The former schemes are the earlier ones, and are dealt with in the next section. Some critical remarks on these schemes follow in section 9. The chapter concludes with some preliminary remarks on the schemes of linear regression.

## 8. A scheme of hidden periodicities.

A simple approach of hidden periodicities is to regard an observational time series (1) as additively built up by a sum of harmonics and a random sample of a certain aleatory variable. Thus, denoting by

$$(38) \quad \xi(1), \xi(2), \dots, \xi(n)$$

the hypothetical random variables corresponding to the observational set (1), we have in this case

$$(39) \quad \xi(t) = y(t) + \eta(t),$$

where  $y(t)$  is of type (17), and  $\eta(t)$  is a set of independent random variables as used in the previous section.

A chief problem when applying a scheme of hidden periodicities is to perform a separation of the functional and the random components. Since we are concerned with the case when  $y(t)$  is a composed harmonic (17), the variate difference method, and other methods based on the assumption that  $y(t)$  reduces to a polynomial or to a trend function, fall outside of the program (cf. p. 1).

The principal method for the search of harmonic components in a given time series (1) is the SCHUSTER periodogram method described in section 5. The test problem concerning the significance of the ordinates in the empirical periodogram was already mentioned in section 7.

In respect to the OPPENHEIM-BRUNS method for separating harmonic components (cf. section 5), it has been emphasized by J. I. CRAIG (1916) that the method fails when a random component is superposed on the harmonics. This disturbing effect of the random error will be called the »CRAIG effect«. As the OPPENHEIM-BRUNS method is parallel to a method of importance in the sequel, and as the CRAIG effect — possibly because of the sketchy character of Mr. CRAIG's paper — seems to have been overlooked in later literature, the point in question will be taken up in a separate discussion. This is done in section 28.

The problem of separating the functional and probabilistic elements being, of course, to a considerable extent indeterminate, even rough methods for the search of periodicities may be of interest. A proximate method, which is of particular relevance because it is both convenient and capable of delivering more general periodic components than harmonic functions, is based on the well-known BUYS-BALLOT table (see e. g. K. STUMPF (1927), p. 100):

$$(40) \quad \left\{ \begin{array}{ccccccc} \bar{\xi}_1, & \bar{\xi}_2, & & \dots & \dots & \dots & \bar{\xi}_p \\ \bar{\xi}_{p+1}, & \bar{\xi}_{p+2}, & & \dots & \dots & \dots & \bar{\xi}_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \bar{\xi}_{(k-1)p+1}, & \bar{\xi}_{(k-1)p+2}, & & \dots & \dots & \dots & \bar{\xi}_{kp} \\ \bar{\xi}_{kp+1}, & \bar{\xi}_{kp+2}, & \dots, & \bar{\xi}_n \end{array} \right.$$
  

$$(41) \quad \bar{m}_1, \quad \bar{m}_2, \quad \dots, \quad \bar{m}_{n-kp}, \quad \dots, \quad \bar{m}_p.$$

Here  $p$  is an integer,  $k$  stands for the greatest integer which is less than  $n/p$ , and  $\bar{m}_i$  is the arithmetical mean of the  $i$ :th col-

umn. Denoting by  $k_i$  the number of elements in the  $i$ :th column, we have

$$(42) \quad \begin{cases} k_i = k + 1 & \text{for } i \leq n - kp, \\ k_i = k & \text{for } i > n - kp. \end{cases}$$

The leading idea of the method simply is that if the series  $\xi_t$  contains a component  $y(t)$  with period  $p$ , the values of  $y(t)$  for  $t = 1, 2, \dots, p$  are approximately given by the means  $\bar{m}_1, \bar{m}_2, \dots, \bar{m}_p$ .

While the arrangement (40) was used even before C. H. D. BUYS-BALLOT (1847), the method was developed in detail by B. STEWART and W. DODGSON (1879) and others (cf. H. BURKHARDT (1904) p. 679 f.). A sharpening of the method by means of a periodogram construction, generalizing that of A. SCHUSTER, is due to E. T. WHITTAKER (1911). In the WHITTAKER periodogram,  $p$  is taken for abscissa, and the ordinate in  $p$  equals the (weighted) variance in the series  $\bar{m}_i$  divided by the variance  $\bar{D}^2(\xi)$  in the series  $\xi_t$ .

The connexion between the two periodogram constructions is interesting. Considering the SCHUSTER periodogram,  $\frac{1}{2} \bar{C}^2(\lambda)$  is well-known to be approximated by the maximum value for varying  $A$  and  $B$  of the expression

$$(43) \quad \bar{D}^2(\xi) - \frac{1}{n} \cdot \sum_{t=1}^n (\xi_t - \bar{m} - A \cdot \cos \lambda t - B \cdot \sin \lambda t)^2.$$

On the other hand, Professor H. CRAMÉR, in his Course in 1933, showed that in the WHITTAKER periodogram the ordinate equals, apart from the constant denominator  $\bar{D}^2(\xi)$ , the maximum of an expression generalizing (43), *viz.*

$$(44) \quad \bar{D}^2(\xi) - \frac{1}{n} \cdot \sum_{t=1}^n [\xi_t - \bar{m} - y_p(t)]^2.$$

This expression should be maximized under the condition that  $y_p(t)$  be a function of integral period  $p$ , but arbitrary for the rest.<sup>1</sup>

With the notations used in (40), the maximum value of (44) equals  $\frac{1}{n} \cdot \sum_{i=1}^p k_i \cdot (\bar{m}_i - \bar{m})^2$ , which for the ordinates in the WHITTAKER periodogram gives

<sup>1</sup> The interpretation of the generalized periodogram as a correlation ratio is incorrect (cf. E. T. WHITTAKER and G. ROBINSON (1926), p. 345 f.).

$$(45) \quad \sum_{i=1}^p k_i (\bar{m}_i - \bar{m})^2 / n \cdot \bar{D}^2(\bar{\xi}).$$

In case  $n$  is a multiple of  $p$ , (45) reduces to

$$\sum_{i=1}^p (\bar{m}_i - \bar{m})^2 / p \cdot \bar{D}^2(\bar{\xi}) = \bar{D}^2(\bar{m}_i) / \bar{D}^2(\bar{\xi}).$$

In his 1933 Course, Prof. H. CRAMÉR also delivered an expectancy theory of the WHITTAKER periodogram. He showed, i. a., that if the given series is a random sample of a normally distributed variable, then  $\bar{D}^2(\bar{m}_i)$  follows the well-known STUDENT's distribution.

It should be observed that as a principle the SCHUSTER and the WHITTAKER periodograms are of equal efficiency — if the former discovers periods in a given observational series, then the latter will give positive results, and *vice versa*.

## 9. On the criticism of the scheme of hidden periodicities.

While the hypothesis of hidden periodicities has proved very fruitful in many fields of scientific research, many applications early met with a serious criticism (cf. H. BURKHARDT (1904) p. 685). The essential point of criticism bears upon the postulated strict periodicity of the individual functional components, and it has been maintained that this rigidity in the periodicity often has no empirical correspondence.

The serial and autocorrelation coefficients disclose an interesting aspect of the above critical argument. If the functional element  $y(t)$  in (39) is a composed harmonic (17), the autocorrelation coefficients  $r_k$  are for  $k \neq 0$  given by

$$(46) \quad r_k = \sum_{i=1}^s C_i^2 \cdot \cos \lambda_i k / (2 D^2(\eta) + \sum_1^s C_i^2).$$

The hypothesis of hidden periodicities assumes, therefore, that, for  $k \neq 0$ , the autocorrelation coefficient  $r_k$ , too, is a function of the type (17), i. e., a composed harmonic. According to the relation (19), the hypothesis thus implies that there exist arbitrarily large  $k$ -values such that  $r_k$  is approximately given by the value taken by (46) for  $k=0$ , viz.

$$(47) \quad \sum_1^s C_i^2 / (2 D^2(\eta) + \sum_1^s C_i^2).$$

This implication may, as a matter of fact, be used as a criterion of the applicability of the hypothesis of hidden periodicities. Thus, even though a given time series clearly shows a cyclical character, but the serial coefficients are gradually vanishing, then the scheme of hidden periodicities is no adequate approach.

It seems plausible that in a good many oscillatory phenomena the serial coefficients actually are gradually vanishing. The above criterion shows that in such cases a periodogram analysis would give negative results. The table of serial coefficients in air pressure material from *Port Darwin* analyzed by SIR G. WALKER ((1931), p. 528) may be referred to for illustration. The graph of serial coefficients presents damped oscillations of a period of about three years.

It follows from the above that in a descriptive analysis of a time series the serial coefficients of G. U. YULE are of fundamental importance. J. BARTELS (1935) has recently given another method of descriptive analysis, consisting in a generalization of the BUYS-BALLOT table, and directly constructed as a criterion of the applicability of the hypothesis of hidden periodicities. J. BARTELS forms the BUYS-BALLOT table for successively extended sections of the given series. Let  $\theta(k, \nu)$  stand for the  $\nu$ :th section consisting of  $k$  consecutive rows in (40), and let  $q$  be the number of such sections. Let  $\bar{D}_p^2(k, \nu)$  be the variance of the column averages  $\bar{m}_i(k, \nu, p)$  in the section  $\theta(k, \nu)$ , and let their arithmetical mean in respect of  $\nu$  be  $\bar{J}_p(k)$ ,

$$\bar{J}_p(k) = \frac{1}{q} \cdot \sum_{\nu=1}^q \bar{D}_p^2(k, \nu).$$

For the expression

$$(48) \quad \bar{B}_p(k) = k \cdot \bar{J}_p(k) / \bar{J}_p(1),$$

regarded as a function of  $k$ , J. BARTELS proposes the term *persistence characteristic* because of the following observations.

Let the series  $\xi_i$  under investigation contain a component of persistent period, i.e. a strictly periodic component. If the length of the period equals  $p$ , and if  $k$  is large, each of the  $q$  sets  $\bar{m}_i(k, \nu, p)$ , considering the variation with  $i$ , will nearly reproduce the periodic

component. Consequently,  $\bar{J}_p(k)$  will tend to a positive constant as  $k \rightarrow \infty$ , and  $\bar{B}_p(k)$  will therefore increase nearly proportionately with  $k$ . — On the other hand, if the time series is purely random,  $\bar{B}_p(k)$  will show no tendency to vary with  $k$ , but will remain on the unity level. — Concerning intermediate cases, J. BARTELS remarks that  $\bar{B}_p(k)$  may tend to an asymptote above unity. Then a quasi-persistency is present, a tendency of adjacent rows in the BUYS-BALLOT table to show a certain resemblance, a resemblance which will fade away as more distant rows are compared.

The persistency characteristics computed by J. BARTELS ((1935) p. 519 f.) suggest persistency in statistical data concerning a) the half year period in the international index of terrestrial magnetism, b) the 24 hour wave in air pressure in *Potsdam*, c) the 12 hour component of the *Batavia* temperature. On the other hand, quasi-persistency is suggested in a) the 27 day component in terrestrial magnetism, b) the 24 hour wave in the magnetic east component in *Batavia*.

In economics, the classical periodogram analysis has repeatedly been tried on business cycle material. The negative results support an opinion which has been maintained also on logical-theoretical grounds, and which now seems predominant, *viz.* that the hypothesis of hidden periodicities is inadequate in business cycle theory.

In cases like those mentioned above, where the approach of rigid periodicity fails, the schemes of linear regression announced in section 7 seem to form a natural and interesting substitute for the scheme of hidden periodicities. Reference to previous results concerning the schemes of linear regression being given later, when dealing systematically with these schemes, the concluding section in this survey will enter into detail only as to the earliest papers on the schemes in question.

#### 10. Remarks on the schemes of linear regression.

In the hypothesis of hidden periodicities, there is assumed a far-going interdependence between the elements of the given time series; leaving the random component out of the question, the interdependence is assumed to be purely functional. The schemes of linear regression assume as to adjacent elements an interdependence only in the sense of the theory of probability.

In an interesting study on the variate difference method, G. U. YULE (1921) considers the autocorrelation in a series consisting of iterated differences, say of order  $m$ , obtained from a purely random series. Since the autocorrelation coefficients are

$$(49) \quad r_k = (-1)^k \cdot \frac{m(m-1)\cdots(m-k+1)}{(m+1)\cdots(m+k)}; \quad k > 0,$$

the series of differences must present an oscillatory character, a feature increasing in evidence with  $m$ . In other words, we are concerned with a primary sequence of random variables, say ...,  $\eta(t-1)$ ,  $\eta(t)$ ,  $\eta(t+1)$ , ..., which by hypothesis are independent, and have identical distribution functions; on this basis a secondary series, say ...,  $\xi(t-1)$ ,  $\xi(t)$ ,  $\xi(t+1)$ , ..., is defined by means of a moving linear operation of the type

$$(50) \quad \xi(t) = b_0 \eta(t) + b_1 \eta(t-1) + \cdots + b_h \eta(t-h).$$

The approach thus defined will in the sequel be called the scheme of *moving averages*.

Another particular case of moving average (50) is studied by E. SLUTSKY (1927), who forms the secondary, intercorrelated series from the primary, purely random series by  $n$  iterated summations by two, followed by the forming of  $m$ :th differences. Holding  $m/n$  constant, E. SLUTSKY shows that, with probability 1, an arbitrarily fixed section of the difference series will tend to a sine curve as  $n \rightarrow \infty$ . This result is given as an application of a general theorem proved in the same paper and discussed in section 16.

Stochastical interdependence of the type (50) is a particular case of linear regression. Letting the auxiliary variables  $\eta(t)$  be the same, another type of linear regression is indicated by the following implicit definition of the intercorrelated variables  $\xi(t)$ ,

$$(51) \quad \xi(t) + a_1 \xi(t-1) + \cdots + a_h \xi(t-h) = \eta(t).$$

The approach (51) was introduced in an heuristic manner in an important memoir by G. U. YULE (1927). The fundamental difference between the scheme of hidden periodicities and the scheme (51), which is said by YULE to define a »disturbed harmonic»  $\xi(t)$ , is clearly brought out. In (39) the random elements  $\eta(t)$  »do not in any way disturb the steady course of the underlying periodic function or functions» (p. 268). On the other hand, regarding (51) YULE

(p. 294) states that a principal feature of a disturbed periodic movement is »*a continual change of amplitude and shift of phase*».

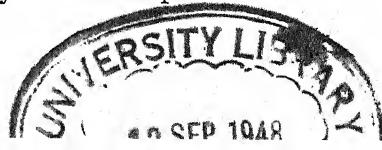
In the same paper G. U. YULE applies, with success, his hypothesis to empirical data, *viz.* A. WOLFER's sunspot numbers. Further, he makes the following general statement concerning the scope of the scheme (51): »*Disturbance will always arise if the value of the variable is affected by external circumstance and the oscillatory variation with time is wholly or partly self-determined, owing to the value of the variable at any one time being a function of the immediately preceding values. Disturbance, as it seems to me, can only be excluded if either (1) the variable is quite unaffected by external circumstance, or (2) we are dealing with a forced vibration and the external circumstances producing this forced vibration are themselves undisturbed*

In order to attain conformity in terminology, the approach (51) will in the sequel be descriptively called the scheme of (*linear*) autoregression. G. U. YULE (1927) restricts himself to the cases  $h \leq 4$ . General autoregression as implicitly defined by (51) was dealt with by SIR G. WALKER (1931), whose applications to the air pressure data mentioned in section 9 gave positive results.

As shown in detail by E. SLUTSKY (see e.g. (1937)), even a scheme of moving averages (50) may present waves of shifting amplitude and phase. Thus, both schemes of linear regression are of interest to the theory of those oscillatory phenomena for which the hypothesis of hidden periodicities proves inadequate. The investigations referred to in section 9 show that there are many central phenomena of this kind.

While the schemes of linear regression thus form a type of hypothesis of the greatest importance, the development of the subject is still little advanced, both as to the theory and the application of the schemes. For instance, earlier definitions concerning the scheme of autoregression are incomplete. One of the chief purposes of the present volume is to give some contributions for completion in these respects. It also aims at bringing the schemes into place in the theory of probability, thereby uniting the rather isolated results hitherto reached.

In the theory of probability, the schemes of linear regression fall under the heading of the discrete stationary random process as defined by A. KHINTCHINE (see (1932) and (1933)). As a matter of fact, the concept of stationary random process is extremely general, and the restrictions involved are only those indispensable



in hypotheses concerning stationary phenomena (cf. p. 3). It is therefore but natural that the scheme of hidden periodicities, after a slight change in the interpretation, will also be found to form a special discrete stationary process (see section 15  $\Omega$ ). Accordingly, the theoretical developments start with a chapter on the discrete random process. This analysis will, i.a., deepen the insight into the nature of the schemes of linear regression, which are dealt with in Chapter III. Chapter IV contains some applications of different hypotheses to observational time series.

In Chapter IV, the serial coefficients of the observational data play a fundamental role. The question concerning the quantitative significance of such correlation coefficients is taken up in Appendix B. Following an argument presented in a preliminary note (see H. WOLD (1936)), it is shown that these coefficients, and likewise all correlation coefficients obtained from a broad class of time and spatial series, have their quantitative significance conditioned by the size of the statistical masses to which the data refer.

## CHAPTER II.

### On the theory of the discrete stationary random process.

#### 11. Definition of the stationary processes.

In the theoretical analysis of a statistical time series, we may distinguish between functional and probabilistic approaches. In the former, the time series is represented by a function of time which in the general case is univalent, but otherwise unconditioned. In the latter, the most general approach is the unconditioned random process. From a purely mathematical viewpoint, the random process is a random variable in an infinite number of dimensions. Denoting by  $\{t\}$  the set of time points  $t$  in which the phenomenon changing with time is studied, each element in  $\{t\}$  corresponds to one dimension in the random variable.

Let  $\{t\}$  stand for a set of values taken on by a real parameter  $t$ , which will be spoken of as representing time, and let one random variable  $\xi(t)$  correspond to each time point  $t$  in  $\{t\}$ . Denoting the given set of random variables by  $\{\xi(t)\}$ , let it be assumed that the following conditions are satisfied.

(A). Choosing arbitrarily a sub-set in  $\{t\}$ , say  $(t) = (t_1, \dots, t_n)$ , the combined variable  $\xi(t_1, \dots, t_n) = [\xi(t_1), \dots, \xi(t_n)]$  will be well-defined.

Let the distribution function of  $\xi(t_1, \dots, t_n)$  be denoted by  $F(t_1, \dots, t_n; u_1, \dots, u_n)$ , so that

$$(52) \quad F(t_1, \dots, t_n; u_1, \dots, u_n) = P[\xi(t_1) \leq u_1, \xi(t_2) \leq u_2, \dots, \xi(t_n) \leq u_n],$$

and let the sets of distribution functions and probability functions of the variables  $\xi(t_1, \dots, t_n)$  be denoted by  $\{F\}$  and  $\{P\}$  respectively.

(B). Letting  $(t) = (t_1, \dots, t_n)$  be an arbitrary sub-set in  $\{t\}$ , and  $(i_1, \dots, i_n)$  be an arbitrary permutation of the sequence  $(1, 2, \dots, n)$ , the functions  $\{F\}$  will satisfy the following relations identically in  $u_1, \dots, u_n$ ,

$$(53) \quad F(t_{i_1}, t_{i_2}, \dots, t_{i_n}; u_{i_1}, u_{i_2}, \dots, u_{i_n}) = F(t_1, \dots, t_n; u_1, \dots, u_n),$$

$$(54) \quad F(t_1, \dots, t_m; u_1, \dots, u_m) = F(t_1, \dots, t_n; u_1, \dots, u_m, +\infty, \dots, +\infty),$$

where  $m < n$ .

These relations express merely that the probability laws ruling the variables  $\{\xi(t)\}$  must not contradict themselves. Accordingly, these relations will be referred to as the *consistency relations*.

Following A. KOLMOGOROFF ((1931), (1933)), a set  $\{\xi(t)\}$  satisfying the conditions (A) and (B) will be called a *random process*.

According to a fundamental theorem of A. KOLMOGOROFF ((1933), p. 27), a set  $\{F\}$  belonging to a random process  $\{\xi(t)\}$  defines a probability distribution on those sets in a space  $R_t$  of an infinite number of dimensions  $\{t\}$ , which are formed by an enumerable sum of BOREL's cylinder sets in  $R_t$ . For instance, if the sequence  $t, t-1, t-2, \dots$  is contained in the set  $\{t\}$ , the probability  $P[\xi(t) \leq u_0, \xi(t-1) \leq u_1, \xi(t-2) \leq u_2, \dots]$  will exist for any real sequence  $u_0, u_1, u_2, \dots$

In order to define stationarity, we must consider arbitrary translations within the set  $\{t\}$ . In doing this, we can assume that  $\{t\}$  either consists of all real  $t$ -values, or is formed by an unbroken sequence of equidistant values, say  $\dots, -1, 0, 1, 2, \dots$ . In any case, a random process  $\{\xi(t)\}$  as defined by a set  $\{F\}$  is termed *stationary* in the sense of A. KHINTCHINE ((1932)—(1934)), if for an arbitrary sub-set  $(t) = (t_1, \dots, t_n)$  in  $\{t\}$  the relation

$$(55) \quad F(t_1+t, t_2+t, \dots, t_n+t; u_1, \dots, u_n) = F(t_1, t_2, \dots, t_n; u_1, \dots, u_n)$$

is identically satisfied in  $u_1, \dots, u_n$  and in  $t$ . Again following KHINTCHINE, the process will be called *discrete*, if  $t$  is restricted to a sequence of equidistant values, *continuous* if  $t$  is arbitrary.

According to the interpretation indicated in section 2, the variables  $\xi(t)$  considered in the above definitions may be taken to be multi-dimensional. Thus, just as in the case of ordinary random variables, a  $k$ -dimensional random process  $\{\xi(t)\}$  may be looked upon as obtained by combining  $k$  one-dimensional processes, say  $\{\xi^{(1)}(t)\}, \dots, \{\xi^{(k)}(t)\}$ . Now, in studying simultaneously a group of one-dimensional processes, we shall always assume that an arbitrary finite sub-group, say  $\{\xi^{(i_1)}(t)\}, \dots, \{\xi^{(i_k)}(t)\}$ , can be combined into a  $k$ -dimensional random process. For instance, considering an infinite sequence  $\{\xi^{(1)}(t)\}, \{\xi^{(2)}(t)\}, \dots$ , this assumption may be expressed as follows.

Taking out arbitrarily a group  $\{\xi^{(i_1)}(t)\}, \dots, \{\xi^{(i_k)}(t)\}$ , fixing arbitrarily a set of time points  $t_1, \dots, t_n$ , and a double real sequence  $u_s^{(r)}$ , where  $r = 1, 2, \dots, k; s = 1, 2, \dots, n$ , we shall assume that the probability

$$P[\xi^{(i_r)}(t_s) \leq u_s^{(r)}; r = 1, \dots, k; s = 1, \dots, n]$$

will exist; further, we shall assume that these probabilities will satisfy all consistency relations of type (53—54); in case of stationarity we shall also assume that all relations of type (55) will hold.

Considering the variables  $\xi(t)$  which constitute a discrete random process  $\{\xi(t)\}$ , let us choose  $t$  arbitrarily, and form the sequence  $\xi(t), \xi(t-1), \xi(t-2)$ , etc. In conformity with the above, this sequence will be denoted  $\xi(t, t-1, t-2, \dots)$  when dealt with as a random variable in an infinite number of dimensions. Such variables being used frequently, we shall adopt the short notation  $\xi[t] = \xi(t, t-1, t-2, \dots)$ . In case the process  $\{\xi(t)\}$  is stationary, it is evident that the probability distribution of  $\xi[t]$  is independent of  $t$ , and — what is likewise of importance — that the distribution of  $\{\xi(t)\}$  will be uniquely determined by that of  $\xi[t]$ .

Expectations derived from the distribution functions  $\{F\}$  determining a stationary process  $\{\xi(t)\}$  will be called *characteristics* of the process. The characteristics are, of course, independent of  $t$ . Further, considering the distribution functions  $F(t; u)$  in the set  $\{F\}$ , these will be independent of  $t$ . The function of  $u$  thus obtained will be termed the *principal* distribution function of the process considered. By definition, the *mean* ( $m$ ), the *dispersion* ( $D$ ), etc., of a one-dimensional stationary process are given by the corresponding characteristics as obtained from the principal distribution function (cf. (5) and (6)).

If the dispersion of a one-dimensional stationary process is finite, the *automoments* of second order as defined by (cf. (10) and (11))

$$\nu_2^{(k)} = E[\xi(t) \cdot \xi(t+k)] = \int_{R_2} u v \cdot du, v F(t, t+k; u, v) = \nu_2^{(-k)}$$

will be finite. The characteristics mentioned determine the *auto-correlation coefficients* belonging to the stationary process  $\{\xi(t)\}$  considered (cf. (12)),

$$(56) \quad r_k = r_k(\xi) = (\nu_2^{(k)} - m^2)/D^2 = r_{-k}$$

If  $r_k(\xi) = 0$  for all  $k \neq 0$ , the process  $\{\xi(t)\}$  will be termed *non-autocorrelated*.

A. KHINTCHINE (1932) gives also a more embracing definition of the stationary process, requiring only that the characteristics  $m$ ,  $D$  and  $\nu^{(k)}$  shall be independent of  $t$ . This case will be referred to as the *generalized* stationary process.

Let  $\{\xi(t)\}$  be a stationary process, and consider the variables  $\xi(t_1, \dots, t_m)$  and  $\xi(t_1, \dots, t_n)$  which refer to the time points  $(t_1, \dots, t_m, \dots, t_n)$ . We shall sometimes have to regard the process  $\{\xi(t)\}$  in the set  $(t_{m+1}, \dots, t_n)$  as conditioned by the behaviour of  $\{\xi(t)\}$  in the time points  $(t_1, \dots, t_m)$ . Following a familiar terminology, we shall then speak of the variable  $\xi(t_{m+1}, \dots, t_n)$  as being *conditioned* by  $\xi(t_1, \dots, t_m)$ . Indicating conditionality by an index  $C$ , and denoting by  $C'$  the condition obtained from  $C$  by replacing throughout  $t_i$  by  $t_i + t$ , it is evident that for an arbitrarily fixed  $t$  the two conditioned variables  $\xi_C(t_{m+1}, \dots, t_n)$  and  $\xi_{C'}(t_{m+1} + t, \dots, t_n + t)$  will have identical distribution functions if the process  $\{\xi(t)\}$  is stationary. The reader is referred to A. KOLMOGOROFF (1933), Chapter V for the fundamentals concerning conditioned variables and distributions.

Generally speaking, a characteristic of a conditioned variable depends on the conditioning variable, and will be called a *conditioned characteristic*. Since such a characteristic forms a function of a random variable, it constitutes in itself a random variable. For instance, considering a one-dimensional process  $\{\xi(t)\}$ , and taking as before  $\xi(t_1, \dots, t_m)$  to be the conditioning variable, the conditioned expectation of  $\xi(t_{m+1})$  is, by definition, the expectation of  $\xi_C(t_{m+1})$ . Denoting this expectation by  $E_C[\xi(t_{m+1})]$ , a general formula gives (see A. KOLMOGOROFF (1933), p. 47)

$$(57) \quad E[E_C[\xi(t_{m+1})]] = E[\xi(t_{m+1})] = m(\xi).$$

*From now on, when not explicitly stated otherwise, the random processes dealt with are tacitly understood (A) to be one-dimensional, (B) to be discrete, and defined for integral time points, (C) to have a finite dispersion.* Of course, the confinement to integral time points instead of a general equidistant sequence, say

$$\dots, t_0 - a, t_0, t_0 + a, t_0 + 2a, \dots$$

involves no restriction as to the generality of the theory of the discrete process.

## 12. A theorem of A. KHINTCHINE.

As far as I am aware, the only earlier investigation of the general discrete stationary process is that of A. KHINTCHINE ((1932), (1933)) already referred to. Though the problems dealt with in the sequel lie along entirely different lines, the principal theorem of KHINTCHINE will be quoted in full because it discloses a fundamental property of the stationary process. The theorem in question states that the stationary process is subjected to the law of great numbers, viz. in the following sense:

*Let  $\xi(t)$ ,  $\xi(t-1)$ , ...,  $\xi(t-n+1)$  be a finite sequence of variables connected with a discrete stationary process  $\{\xi(t)\}$  with finite dispersion, and put  $\Sigma_n = \frac{1}{n} \cdot \sum_{i=0}^{n-1} \xi(t-i)$ . The dispersion of the difference  $\Sigma_n - \Sigma_m$  then tends to zero when  $n \rightarrow \infty$  and  $m \rightarrow \infty$ . Further, processes may be constructed so that the asymptotical decrease is arbitrarily slow.*

In view of the applications — in particular certain questions concerning ergodic hypotheses — it is an interesting problem whether in the sums  $\Sigma$  the sequence  $i = 0, 1, 2, \dots$  can be replaced by the sequence  $i_0 < i_1 < i_2 < \dots$  A short reflection on the singular processes as defined and exemplified in sections 14—16 shows that such a general sequence is not allowed.

The problems dealt with in the sequel have their points of connection with certain investigations on the continuous process and on special types of the discrete process. These earlier investigations will be referred to in the course of the analysis.

The concept of stationary process as introduced by A. KHINTCHINE is extremely general. As a scheme for the analysis of time series it will be found to embrace all the schemes mentioned in the survey given in Chapter I. For the sake of concreteness it will be of interest, before passing to the general theoretical developments, to show in detail how these may be obtained by suitable specializations. To this end a preparatory analysis of the general process will be useful. Accordingly, the next two sections will be reserved for some groundwork concerning operation with the discrete stationary processes.

## 13. Some fundamental operations with random processes.

In this section it will be shown that certain familiar operations with ordinary random variables can also be performed with random

processes. In discussing the situation, no restrictions will be laid on the processes considered. Generally speaking, the operations will give rise to new random processes. Moreover, if the processes dealt with are stationary, the resulting processes will be found to be stationary. It will be sufficient for our purpose to consider functions of random processes, and the forming of limit processes in convergent sequences.

Denoting by  $\xi$  a random variable in a  $k$ -dimensional space  $R_k$ , and by  $f[x]$  a function which is finite and BOREL-measurable in  $R_k$ , and whose values are lying in a space  $R_p$ , it is known that  $f[\xi]$  will be a well-defined random variable in  $R_p$  (see e. g. H. CRAMÉR (1937), p. 12 f). Let us next consider a combined variable  $[\xi^{(1)}, \dots, \xi^{(n)}]$  consisting of random variables  $\xi^{(i)}$  in  $R_k$ . Forming the variables  $f[\xi^{(i)}]$ , it is evident that these in the same way may be combined to a random variable  $[f[\xi^{(1)}], \dots, f[\xi^{(n)}]]$  (cf. also p. 10).

Thus prepared, let  $f[x]$  remain the same, and consider the variables  $\xi(t_1, \dots, t_n) = [\xi(t_1), \dots, \xi(t_n)]$  which constitute a random process  $\{\xi(t)\}$ . According to the above, the variables  $[f[\xi(t_1)], \dots, f[\xi(t_n)]]$  will be well-defined. Denoting by  $\{F^*\}$  the set of distribution functions of these variables, it is also evident that the functions  $F^*$  satisfy all consistency relations of type (53—54). Further, if  $\{\xi(t)\}$  is stationary, the functions  $F^*$  also will satisfy (55). The variables  $[f[\xi(t_1)], \dots, f[\xi(t_n)]]$  will thus constitute a random process, and if  $\{\xi(t)\}$  is stationary, the process obtained will also be stationary. The resulting process will be said to be a function of the process  $\{\xi(t)\}$ , and be denoted by  $\{f[\xi(t)]\}$ . The variables of type  $[f[\xi(t)], f[\xi(t-1)], \dots, f[\xi(t-n)]]$  will be denoted by  $f[\xi(t, t-1, \dots, t-n)]$ .

In particular, considering a random process  $\{\xi(t)\}$  obtained by combining  $k$  one-dimensional processes, say  $\{\xi^{(1)}(t)\}, \dots, \{\xi^{(k)}(t)\}$ , let us take  $f$  to be linear. The operation will then give rise to a sum process of type  $\{a_1 \xi^{(1)}(t) + \dots + a_k \xi^{(k)}(t)\}$ . Denoting this sum by  $\{\zeta_k(t)\}$ , we shall write

$$\{\zeta_k(t)\} = a_1 \{\xi^{(1)}(t)\} + \dots + a_k \{\xi^{(k)}(t)\}.$$

According to the above, we have

$$(58) \quad \zeta_k(t_1, \dots, t_n) = a_1 \xi^{(1)}(t_1, \dots, t_n) + \dots + a_k \xi^{(k)}(t_1, \dots, t_n),$$

$$\zeta_k[t] = a_1 \xi^{(1)}[t] + a_2 \xi^{(2)}[t] + \dots + a_k \xi^{(k)}[t].$$

Most of the functional operations dealt with in the sequel are of this simple kind. As to the processes  $\{\xi^{(1)}(t)\}, \dots, \{\xi^{(k)}(t)\}$  combined, these, too, are for the most part of a simple structure. We shall next present two type cases.

Let  $\{\xi^{(1)}(t)\}, \dots, \{\xi^{(k)}(t)\}$  represent a set of random processes and let a set of time points  $(t) = (t_1, \dots, t_n)$  and  $k$  sets of real numbers  $(u^{(s)}) = (u_1^{(s)}, \dots, u_n^{(s)})$ ;  $s = 1, \dots, k$ , be chosen arbitrarily. Then the processes  $\{\xi^{(s)}\}$  will be called *independent* if the following relation is satisfied

$$\begin{aligned} P[\xi^{(1)}(t_1) \leq u_1^{(1)}, \dots, \xi^{(1)}(t_n) \leq u_n^{(1)}; \dots; \xi^{(k)}(t_1) \leq u_1^{(k)}, \dots, \xi^{(k)}(t_n) \leq u_n^{(k)}] = \\ = P[\xi^{(1)}(t_1) \leq u_1^{(1)}, \dots, \xi^{(1)}(t_n) \leq u_n^{(1)}] \dots P[\xi^{(k)}(t_1) \leq u_1^{(k)}, \dots, \xi^{(k)}(t_n) \leq u_n^{(k)}]. \end{aligned}$$

Similarly, a sequence  $\{\xi^{(1)}(t)\}, \{\xi^{(2)}(t)\}, \dots$  will be said to consist of independent processes if the processes in every finite subset  $\{\xi^{(i_1)}(t)\}, \dots, \{\xi^{(i_k)}(t)\}$  are independent.

Now, let it be assumed that the independent processes  $\{\xi^{(s)}(t)\}$  are stationary, and have finite dispersions  $D(\xi^{(s)})$ . Then the sum process  $\{\zeta_k(t)\}$  as defined by (58) will be stationary, and the expectation, the dispersion, and the autocorrelation coefficients of  $\{\zeta_k(t)\}$  will exist. We have  $E[\zeta_k] = a_1 E[\xi^{(1)}] + a_2 E[\xi^{(2)}] + \dots + a_k E[\xi^{(k)}]$ , and, as is readily verified,

$$(59) \quad \begin{cases} D^2(\zeta_k) = a_1^2 D^2(\xi^{(1)}) + a_2^2 D^2(\xi^{(2)}) + \dots + a_k^2 D^2(\xi^{(k)}), \\ r_p(\zeta_k) = a_1^2 \frac{D^2(\xi^{(1)})}{D^2(\zeta_k)} \cdot r_p(\xi^{(1)}) + \dots + a_k^2 \frac{D^2(\xi^{(k)})}{D^2(\zeta_k)} \cdot r_p(\xi^{(k)}). \end{cases}$$

Of course, the two latter relations depend on the identities

$$(60) \quad r(\xi^{(r)}(t \pm p); \xi^{(s)}(t \pm q)) = 0, \quad p \geq 0, q \geq 0,$$

where  $r$  and  $s$  are arbitrary. Having stated this,  $\{\xi^{(r)}\}$  and  $\{\xi^{(s)}\}$  will be termed *non-correlated* or *uncorrelated* if the relations (60) are satisfied. Evidently, the relations (59) hold under the broader assumption that any two processes  $\{\xi^{(r)}\}$  and  $\{\xi^{(s)}\}$  are non-correlated.

In order to define the second type case, let us consider the variables  $\xi(t_1, \dots, t_n)$  which constitute a random process  $\{\xi(t)\}$ . Choosing arbitrarily an integer  $k$ , let us form a second type of variable, say  $\xi^*(t_1, \dots, t_n)$ , by taking  $\xi^*(t_1, \dots, t_n) = \xi(t_1 + k, \dots, t_n + k)$ . Evidently,

these variables  $\xi^*$  constitute a random process. Now, denoting this process by  $\{\xi(t+k)\}$ , a short reflection shows that we may combine the processes  $\{\xi(t)\}$  and  $\{\xi(t+k)\}$ . In fact, the distribution functions ruling the simultaneous behaviour of  $\{\xi(t)\}$  and  $\{\xi(t+k)\}$  are uniquely determined by the distribution functions  $\{F\}$  ruling the process  $\{\xi(t)\}$ , and it is further evident that the resulting distribution functions will satisfy all consistency relations (53—54). Moreover, if  $\{\xi(t)\}$  is stationary, the combined process will be stationary.

The arguments being perfectly general, we can form the processes  $\{\xi(t)\}$ ,  $\{\xi(t-1)\}$ , ...,  $\{\xi(t-h)\}$  and combine them into an  $(h+1)$ -dimensional process. Now, if  $\{\xi(t)\}$  is one-dimensional, we can apply a linear operation of type (58) to the combined process. The result will be a process, say  $\{\zeta(t)\}$ , such that the corresponding variables  $\zeta(t_1, \dots, t_n)$  and  $\zeta[t]$  will satisfy relations like

$$(61) \quad \begin{aligned} \zeta(t_1, \dots, t_n) = & a_0 \xi(t_1, \dots, t_n) + \\ & + a_1 \xi(t_1 - 1, \dots, t_n - 1) + \dots + a_h \xi(t_1 - h, \dots, t_n - h) \end{aligned}$$

$$(62) \quad \zeta[t] = a_0 \xi[t] + a_1 \xi[t-1] + \dots + a_h \xi[t-h].$$

If  $\{\xi(t)\}$  is stationary, then  $\{\zeta(t)\}$  will also be stationary.

The operations considered above may also be applied to observational time series. Letting  $\dots, \xi_{t-1}, \xi_t, \xi_{t+1}, \dots$  represent such a series, and transforming by means of a function  $f[x]$ , the resulting series will read  $\dots, f[\xi_{t-1}], f[\xi_t], f[\xi_{t+1}], \dots$ . If, in particular, every  $\xi_t$  consists of a couple of  $k$  observations, say  $(\xi_t^{(1)}, \dots, \xi_t^{(k)})$ , and if  $f[x]$  is linear, the transformed series, say  $\dots, \bar{\xi}_{t-1}, \bar{\xi}_t, \bar{\xi}_{t+1}, \dots$ , will have for general element

$$\bar{\xi}_t = a_1 \xi_t^{(1)} + a_2 \xi_t^{(2)} + \dots + a_k \xi_t^{(k)}.$$

On the other hand, assuming  $\xi_t$  to be one-dimensional, the transform (61) corresponds simply to a moving linear operation. In this case the general element in the transformed series reads

$$(63) \quad \bar{\xi}_t = a_0 \xi_t + a_1 \xi_{t-1} + \dots + a_h \xi_{t-h}.$$

In the theoretical developments, we shall frequently have to consider sums of type (58) when the number of terms tends to infinity. For use in such connexions, we need a suitable definition of convergence. To this end we shall next extend the concept of con-

vergence in probability, as introduced by F. P. CANTELLI (1916), so as to apply to a sequence of random processes.

A sequence  $\xi^{(1)}, \xi^{(2)}, \dots$  of ordinary random variables is said to converge in probability to a random variable  $\xi$  if for every  $\varepsilon > 0$

$$P[|\xi^{(n)} - \xi| > \varepsilon]$$

tends to zero as  $n \rightarrow \infty$ . A necessary and sufficient condition for such convergence is that, for every  $\varepsilon > 0$ , there exists a number  $n$  such that for an arbitrary  $q > 0$  the following inequality holds (see A. KOLMOGOROFF (1933), p. 32).

$$P[|\xi^{(n+q)} - \xi^{(n)}| > \varepsilon] < \varepsilon.$$

Using the familiar interpretation of  $|x - y|$  as the distance between two points  $x$  and  $y$  in a multi-dimensional space, the definition of convergence also holds in case the variables are  $k$ -dimensional, say  $\xi^{(i)} = [\xi_1^{(i)}, \dots, \xi_k^{(i)}]$ ,  $\xi = [\xi_1, \dots, \xi_k]$ . Of course, an equivalent definition is that for every  $\varepsilon > 0$  the probability

$$P[|\xi_1^{(n)} - \xi_1| < \varepsilon, \dots, |\xi_k^{(n)} - \xi_k| < \varepsilon]$$

tends to unity as  $n \rightarrow \infty$ . Now, considering the elementary inequality of G. BOOLE,

$$(64) \quad P[|\xi_r^{(n)} - \xi_r| > \varepsilon] \leq 1 - P[|\xi_1^{(n)} - \xi_1| \leq \varepsilon, \dots, |\xi_k^{(n)} - \xi_k| \leq \varepsilon] \leq \\ \leq P[|\xi_1^{(n)} - \xi_1| > \varepsilon] + \dots + P[|\xi_k^{(n)} - \xi_k| > \varepsilon],$$

it is evident that a necessary and sufficient condition that  $\xi^{(n)}$  converges in probability to  $\xi$  is that  $\xi_r^{(n)}$  converges in probability to  $\xi_r$  for  $r = 1, \dots, k$  (see F. P. CANTELLI (1916) and E. SLUTSKY (1925)).

Thus prepared, let a sequence of random processes be denoted by

$$(65) \quad \{\xi^{(1)}(t)\}, \{\xi^{(2)}(t)\}, \dots$$

The sequence will be called convergent in probability to a limit process  $\{\xi(t)\}$  if for an arbitrary set  $(t) = (t_1, \dots, t_n)$  the sequence

$$(66) \quad \xi^{(1)}(t_1, \dots, t_n), \xi^{(2)}(t_1, \dots, t_n), \dots$$

be convergent in probability to the limit variable  $\xi(t_1, \dots, t_n)$ .

*Theorem 1.* A necessary and sufficient condition that a sequence (65) of random processes be convergent in probability is that for an arbitrary  $t$  the sequence

$$(67) \quad \xi^{(1)}(t), \xi^{(2)}(t), \dots$$

be convergent in probability. If the sequence (65) is convergent, and if every process  $\{\xi^{(n)}(t)\}$  is stationary, the limit process will be stationary.

The necessity of the condition is implied in the above definition of convergence in probability. Next, let  $(t) = (t_1, \dots, t_n)$  be arbitrarily fixed, and consider the sequence (66). According to the above application of the inequality of G. BOOLE, the convergence of (66) is implied in the convergence of (67) for every  $t$  in the set  $(t_1, \dots, t_n)$ . Having stated this, let the limit variable be denoted by  $\xi(t_1, \dots, t_n)$ , and consider the limit variables belonging to all possible sets  $(t) = (t_1, \dots, t_n)$ . Since the consistency relations of type (53—54) are satisfied for every process in the sequence (65), the same relations must be satisfied in the limit. The variables  $\xi(t_1, \dots, t_n)$  will thus constitute a random process, say  $\{\xi(t)\}$ , and the same argument shows that the limit process  $\{\xi(t)\}$  is stationary if every process in the sequence (65) is stationary.

The following corollary needs no comment.

*Corollary.* Letting  $\{\xi(t)\}$  represent a random process, the sequence (65) converges in probability to  $\{\xi(t)\}$  if, and only if, for an arbitrary  $t$  the sequence (67) converges in probability to  $\xi(t)$ .

In dealing with stationary sequences (65), the theorem proved above will be particularly useful, for the behaviour of (67) in respect of convergence will then be independent of  $t$ . Considering, in particular, the sum process  $\{\zeta_k(t)\}$  defined by (58), a necessary and sufficient condition for convergence in probability as  $k \rightarrow \infty$

is that the sum  $\sum_{i=1}^{\infty} a_i \xi^{(i)}(t)$  is convergent. If convergence takes place,

the sums  $\sum_{i=1}^{\infty} a_i \{\xi^{(i)}(t)\}$  and  $\sum_{i=1}^{\infty} a_i \xi^{(i)}[t]$  will be said to converge.

A sufficient condition of convergence is, of course, that the sum  $\sum_1^{\infty} a_i \cdot E[\xi^{(i)}(t)]$  is convergent, and that the dispersion of  $\sum_{i=n}^{n+p} a_i \xi^{(i)}(t)$

tends to zero uniformly in  $p$  as  $n \rightarrow \infty$ . Moreover, writing  $m_n$  and  $\sigma_n$  for the mean and dispersion of  $\sum_{i=1}^n a_i \cdot \xi^{(i)}(t)$ , it follows readily that if these conditions are satisfied, it would imply a contradiction were not the mean and dispersion of the limit variable given by  $\lim_{n \rightarrow \infty} m_n$  and  $\lim_{n \rightarrow \infty} \sigma_n$  respectively.

As a second application of theorem 1 we make the following observation. Denoting by  $\{\xi(t)\}$  an arbitrary discrete stationary process, let a sequence (65) of processes be defined by

$$\{\xi^{(i)}(t)\} = [\dots, \xi(t+1), \xi(t), \xi(t-1), \dots, \xi(t-i), 0, 0, \dots].$$

Then for every fixed  $(t) = (t_1, \dots, t_n)$  we have  $\lim_{i \rightarrow \infty} \xi^{(i)}(t_1, \dots, t_n) = \xi(t_1, \dots, t_n)$ . It follows that  $\xi^{(i)}(t)$  converges in probability to  $\{\xi(t)\}$ .

Extending a current terminology, two processes, say  $\{\xi^{(1)}(t)\}$  and  $\{\xi^{(2)}(t)\}$ , will be called *equivalent*, if for an arbitrary  $(t) = (t_1, \dots, t_n)$  the two variables  $\xi^{(1)}(t_1, \dots, t_n)$  and  $\xi^{(2)}(t_1, \dots, t_n)$  are equivalent, i. e. if

$$P[\xi^{(1)}(t_1, \dots, t_n) \neq \xi^{(2)}(t_1, \dots, t_n)] = 0.$$

If two variables or processes are equivalent, we shall write  $\xi^{(1)} = \xi^{(2)}$ ,  $\{\xi^{(1)}(t)\} = \{\xi^{(2)}(t)\}$ , etc.

#### 14. On singular stationary processes.

Let  $\xi = [\xi^{(1)}, \dots, \xi^{(n)}]$  represent an  $n$ -dimensional random variable with distribution function  $F(u_1, \dots, u_n)$ . The distribution of  $\xi$  will be called *linearly singular* or, more briefly, *singular*, if there exists a linear function, say  $L[x - m] = a_1(x^{(1)} - m_1) + \dots + a_n(x^{(n)} - m_n)$ , such that

$$(68) \quad P[L[\xi - m] \neq 0] = \\ = P[a_1(\xi^{(1)} - m_1) + \dots + a_n(\xi^{(n)} - m_n) \neq 0] = 0.$$

If (68) is satisfied, the variables  $\xi^{(i)}$  will be said to be connected by the relation  $L[\xi - m] = 0$ . The singularity will be said to be of *rank h*, if there exist  $n-h$ , and only  $n-h$ , independent relations between the variables  $\xi^{(i)}$ , say

$$(69) \quad \begin{cases} a_{1,h+1} \cdot (\xi^{(1)} - m_1) + \cdots + a_{n,h+1} \cdot (\xi^{(n)} - m_n) = 0 \\ \vdots \\ a_{1,n} \cdot (\xi^{(1)} - m_1) + \cdots + a_{n,n} \cdot (\xi^{(n)} - m_n) = 0. \end{cases}$$

It will be relevant to interpret singularity in terms of characteristic functions. Denoting the characteristic function of  $\xi$  by  $f(X_1, \dots, X_n)$ , we have by definition

$$\begin{aligned} f(X_1, \dots, X_n) &= E[e^{i(X_1 \cdot \xi^{(1)} + \cdots + X_n \cdot \xi^{(n)})}] = \\ &= \int_{\mathbb{R}_n^n} e^{i(X_1 \cdot x_1 + \cdots + X_n \cdot x_n)} \cdot dF(x_1, \dots, x_n). \end{aligned}$$

Letting a matrix of real elements be given by

$$(70) \quad \begin{pmatrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}, & a_{n2}, & \dots, & a_{nn} \end{pmatrix}$$

and writing  $(Z) = (Z_1, \dots, Z_n)$  for an auxiliary set of real variables, let the substitution

$$(71) \quad \begin{cases} X_1 = a_{11} \cdot Z_1 + \cdots + a_{1n} \cdot Z_n \\ \vdots \\ X_n = a_{n1} \cdot Z_1 + \cdots + a_{nn} \cdot Z_n \end{cases}$$

transform  $f(X_1, \dots, X_n)$  into  $f^*(Z_1, \dots, Z_n)$ . Considering the variable  $\zeta = (\zeta^{(1)}, \dots, \zeta^{(n)})$  defined by

$$(72) \quad \begin{cases} \zeta^{(1)} = a_{11} \cdot \xi^{(1)} + \cdots + a_{n1} \cdot \xi^{(n)} \\ \vdots \\ \zeta^{(n)} = a_{1n} \cdot \xi^{(1)} + \cdots + a_{nn} \cdot \xi^{(n)}, \end{cases}$$

an elementary transformation shows that

$$(73) \quad f^*(Z_1, \dots, Z_n) = E[e^{i(X_1 \cdot \xi^{(1)} + \cdots + X_n \cdot \xi^{(n)})}] = E[e^{i(Z_1 \cdot \zeta^{(1)} + \cdots + Z_n \cdot \zeta^{(n)})}].$$

Thus  $f^*$  is nothing else than the characteristic function of the composite variable  $\zeta$ . Now, let  $\xi$  be singular of rank  $h$ , say on account of the relations (69), and introduce for a moment the inconsequential assumption  $m_i = 0$ . In such case  $f^*(Z_1, \dots, Z_n)$  will contain at most the variables  $Z_1, \dots, Z_h$ . On the other hand, if the matrix (70) is non-singular the first  $h$  variables must all appear in  $f^*(Z_1, \dots, Z_n)$ . In fact, since the distribution of  $\zeta$  is uniquely determined by its

characteristic function (see e.g. H. CRAMÉR (1937) p. 104), the vanishing of  $Z_m$  in  $f^*$  for a number  $m \leq h$  would imply  $\zeta^{(m)} = 0$ , a linear relation not obtainable from (69).

It will be observed that formula (73) holds for arbitrary values of the real coefficients  $a_{ik}$ . In particular, the characteristic function of the  $h < n$  first variables  $\xi^{(i)}$  as given by (72) will be arrived at by putting  $a_{ik} = 0$  for  $k = h + 1, h + 2, \dots, n$ .

If  $\xi$  is singular on account of a relation of type (68), and if  $E[\xi^{(i)}]$  exists, we shall always assume that

$$(74) \quad m_i = E[\xi^{(i)}],$$

which evidently will involve no loss of generality. Now, calculating  $E[\zeta^{(i)}]$  from (72), and paying regard to (74), we get

$$(75) \quad \begin{cases} m_1^* = a_{11} m_1 + \dots + a_{n1} m_n = E[\zeta^{(1)}] \\ \dots \dots \dots \dots \dots \dots \\ m_n^* = a_{1n} m_1 + \dots + a_{nn} m_n = E[\zeta^{(n)}]. \end{cases}$$

Next, let it be assumed that all the variables  $\xi^{(i)}$  have a finite dispersion. Then the expectation

$$\begin{aligned} E[(X_1(\xi^{(1)} - m_1) + \dots + X_n(\xi^{(n)} - m_n))^2] &= \\ &= \int_{\mathbb{R}^n} [X_1(x_1 - m_1) + \dots + X_n(x_n - m_n)]^2 \cdot dF(x_1, \dots, x_n) \end{aligned}$$

will exist, and represent a non-negative quadratic form in the real variables  $X_i$ , say  $Q(X_1, X_2, \dots, X_n)$ . Writing

$$\mu_{ik} = E[(\xi^{(i)} - m_i)(\xi^{(k)} - m_k)],$$

we have  $\mu_{ik} = \mu_{ki}$ , and

$$(76) \quad Q(X_1, X_2, \dots, X_n) = \sum_{i=1}^n \sum_{k=1}^n \mu_{ik} \cdot X_i X_k.$$

The quadratic form  $Q$  will give information concerning the singularity of the distribution of  $\xi$ . In fact, the rank of the quadratic form equals the rank of the distribution. In other words, by a suitable substitution in  $Q$  of type (71) the number of variables may be brought down to the rank of  $\xi$ , but not further — and vice versa. Denoting the transformed forms by  $Q^*$ , we have also

$$Q^*(Z_1, \dots, Z_n) = E[(Z_1(\zeta^{(1)} - m_1^*) + \dots + Z_n(\zeta^{(n)} - m_n^*))^2],$$

so the forms  $Q^*$  thus are quite analogous to the characteristic functions  $f^*$  in respect of linear substitutions. As the proofs, too, may be given a parallel wording, the verification of the statements made needs no comment.

After these preliminaries, the next step will be to investigate as to stationarity the variables  $\xi(t_1, \dots, t_n)$  connected with a stationary process  $\{\xi(t)\}$ . Until further notice, we shall not assume that  $\{\xi(t)\}$  has a finite dispersion. As before it will be sufficient to consider the variables of type  $\xi(t, t-1, \dots, t-n)$ . Let it first be observed that if  $\xi(t, t-1, t-n)$  is singular on account of the relation

$$(77) \quad \xi(t) - m + a_1 \cdot (\xi(t-1) - m) + \cdots + a_h \cdot (\xi(t-h) - m) = 0,$$

where  $h \leq n$ , then (a) the relation of singularity (77) must hold for every  $t$ , (b) if  $h < n$ , the variable  $\xi(t, t-1, \dots, t-h)$  must also present the same singularity. Further, if  $\xi(t, t-1, \dots, t-h)$  satisfies a second linear relation of type (77), say with coefficients  $(a'_1, \dots, a'_h)$ , we obtain by subtraction a linear relation showing that there is a  $h' < h$  such that  $\xi(t, t-1, \dots, t-h')$  is singular. Thus, when considering a stationary process  $\{\xi(t)\}$ , a number  $h$  will be well-defined by terming  $\{\xi(t)\}$  singular of rank  $h$  if in the sequence

$$\xi(t), \quad \xi(t, t-1), \quad \xi(t, t-1, t-2), \quad \dots, \quad \xi(t, t-1, \dots, t-n), \quad \dots$$

the variable  $\xi(t, t-1, \dots, t-h)$  is the first singular one. Taking (77) for the singularity relation of  $\xi(t, t-1, \dots, t-h)$ , it is readily seen that

- (A) the coefficients  $a_i$  are uniquely determined, and  $a_h \neq 0$ .  
 (B) for all  $k \geq 0$  the  $(k + h + 1)$ -dimensional variable  $\xi(t_0 + k, t_0 + k - 1, \dots, t_0 - h)$  is singular of rank  $h$ , and a system of  $k + 1$  independent relations is given by

Thus, if (77) holds the variable  $\zeta(t_1, \dots, t_n)$  defined by (61), taking  $a_0 = 1$ , will reduce to  $[E[\zeta], \dots, E[\zeta]]$ . Having stated this, the relations of type (78) which follow from (77) in the case of stationarity, may be comprehended in the following identity

$$(79) \quad \xi[t] - m + a_1 \cdot (\xi[t-1] - m) + \dots + a_h (\xi[t-h] - m) = 0.$$

The term *relation of singularity* will be used also for (79).

According to observation (B), a sample value

$$\xi_i(t_0 + k, t_0 + k - 1, \dots, t_0 - h) = [\xi_i(t_0 + k), \xi_i(t_0 + k - 1), \dots, \xi_i(t_0 - h)]$$

will, when regarding  $\xi_i(t_0 + t)$  as a function of  $t$ , with probability 1 satisfy the difference equation (32). Thus, if this difference equation reduces to (30), any sample series

$$[\xi_i(t_0 + k), \xi_i(t_0 + k - 1), \dots, \xi_i(t_0 - h)]$$

will with probability 1 be of type (17), i. e. consist of a number of superposed harmonics. Speaking in this case of a *process of superposed harmonics*, we have the following theorem.

*Theorem 2.* Let  $\{\xi(t)\}$  be a discrete stationary process with autocorrelation coefficients  $r_k$ . If  $\{\xi(t)\}$  is linearly singular,  $\{\xi(t)\}$  is a process of superposed harmonics. A necessary and sufficient condition that  $\{\xi(t)\}$  be linearly singular, say on account of the relation  $L[\xi(t) - m] = 0$  given by (77), is that  $r_k$  satisfies the difference equation  $L[r_k] = 0$ .

Denoting by  $m$  and  $\sigma$  the mean and dispersion of the process considered, the autocorrelation coefficients are given by (cf. (56))

$$(80) \quad r_k = r_{-k} = E[(\xi(t) - m)(\xi(t-k) - m)] / \sigma^2.$$

Thus the quadratic form of type (76) belonging to  $\xi(t, t-1, \dots, t-n)$ , say  $Q_n(X_t, X_{t-1}, \dots, X_{t-n})$ , will be well-defined, and given by

$$Q_n(X_t, \dots, X_{t-n}) = \sigma^2 \cdot \sum_{p=0}^n \sum_{q=0}^n r_{|p-q|} \cdot X_{t-p} X_{t-q}.$$

Since the form  $Q_n$  is non-negative definite, its principal determinant will be non-negative (see e. g. G. KOWALEWSKI (1909), Chapter 12),

$$(81) \quad \mathcal{A}(r, n) = \begin{vmatrix} 1, & r_1, & r_2, & \dots, & r_n \\ r_1, & 1, & r_1, & \dots, & r_{n-1} \\ r_2, & r_1, & 1, & \dots, & r_{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ r_n, & r_{n-1}, & r_{n-2}, & \dots, & 1 \end{vmatrix} \geq 0.$$

The determinants  $\mathcal{A}(r, n)$  defined above will be called the *principal correlation determinants* of the stationary process examined.

Thus prepared, let us begin with proving the second part of the theorem. In the first place, let  $\{\xi(t)\}$  be singular on account of (77), and multiply the left member of this relation by  $\xi(t-k)/\sigma^2$ . Observing that the expectation of the resulting expression is zero, and paying regard to (80), we get

$$(82) \quad L[r_k] = r_k + a_1 r_{k-1} + \dots + a_h r_{k-h} = 0.$$

This relation shows that the condition is necessary.

On the other hand, let the autocorrelation coefficients  $r_t$  satisfy a linear difference equation  $L[r_t] = 0$  of order  $h^*$ . Transforming this equation to the form (32), and reducing to lowest possible order, say  $h$ , let  $L_1[r_t] = 0$  be the result. According to the previous analysis,  $\xi(t)$  will then satisfy no linear difference relation of order  $< h$ . After this remark, let the consecutive rows of  $\mathcal{A}(r, h)$  be denoted by  $q_0, q_1, \dots, q_h$ . From the structure of  $\mathcal{A}(r, h)$  it is evident that these rows are connected by the linear relation  $L_1[q_h] = 0$ . Thus  $\mathcal{A}(r, h)$  equals zero, so the rank of  $\mathcal{A}$  will be  $\leq h$ . Recalling from the theory of quadratic forms that the rank of  $\mathcal{A}(r, n)$  equals the rank of  $Q_n$ , and keeping in mind the identity between the ranks of corresponding distributions and quadratic forms of type (76), it turns out that  $\xi(t, t-1, \dots, t-h)$  is linearly singular. Let the relation of singularity be  $L_2[\xi(t) - m] = 0$ . Now, were not  $L_2 \equiv L_1$ , then, contrarily to the assumption made,  $r_t$  would satisfy a linear difference relation of order  $< h$ . Moreover, according to the construction of  $L_1$  we have  $L[r_k] = L^*[L_1[r_k]]$ , where  $L^*$  is a well-defined linear operation. Thus  $L[\xi(t) - m] = L^*[L_1[\xi(t) - m]] = L^*[0] = 0$ , which proves that the condition is sufficient.

In order to prove the first part of theorem 2, let (77) be the relation of singularity, and let it be assumed that this has already been reduced to the lowest possible order. According to the above,  $r_k$  will then satisfy (82), but no linear equation of lower order.

Writing  $r_k$  on the form (33), this implies that none of the polynomials  $H$  will be vanishing identically. Since, finally, the inequality  $|r_k| \leq 1$  shows that  $r_k$  is uniformly bounded in modulus in  $(-\infty < k < \infty)$ , we conclude from the second remark in section 6 that  $r_k$  can be written on the form (17).

After this analysis, the following theorem will need no comment.

*Theorem 3.* Let  $\{\xi(t)\}$  be a discrete stationary process with principal correlation determinants  $A(r, n)$  given by (81). A necessary and sufficient condition that  $\{\xi(t)\}$  be singular of rank  $h$  is that  $A(r, h)$  be the first vanishing determinant in the sequence  $A(r, 1), A(r, 2), \dots$

A relation of type (77) will be called a *stochastical difference relation of order  $h$*  satisfied by the stationary process  $\{\xi(t)\}$ . The previous analysis shows that a stationary process which has finite dispersion and satisfies (77) will satisfy a difference relation of type

$$(83) \quad A^{2s} \xi(t-s) + h_1 A^{2s-2} \xi(t-s+1) + \dots + h_s [\xi(t) - m] = 0.$$

In the next two sections it will be shown, i. a., that the theorems of the present section are not vacuous, i. e. that there really exist stationary processes having the properties assumed by hypothesis.

### 15. Some type cases of the discrete stationary process.

As mentioned in section 12, it will be shown in the present section that the schemes surveyed in Chapter I may be regarded as special cases of the discrete stationary process. Some other schemes will also be presented, and a few characteristic properties of the different types considered be pointed out. Conditioned variables and expectations, and the operation of addition will be exemplified, and devices given for the construction of model series which follow the different schemes. For further concreteness some model series constructed for the illustration of later results will be furnished.

a. *The purely random process.* This term will in the sequel be used for the purely random scheme touched upon in section 7. The purely random process will, of course, be obtained by taking, in the relation (52) defining the general process,

$$F(t_1, \dots, t_n; u_1, \dots, u_n) = F(u_1) \dots F(u_n),$$

where any distribution function may be chosen for  $F(u)$ . The verification of (53)–(55) is obvious.

When detailed information is required, a purely random process  $\{\xi(t)\}$  defined by a distribution function  $F(u)$  will be denoted by  $\{\xi(t; F)\}$ . It is seen that the defining function  $F(u)$  is identical with the principal distribution function of the process.

The following simple theorem exemplifies the operation of addition of independent processes.

*Theorem 4.* Let  $\{\xi^{(1)}(t; F^{(1)})\}$ ,  $\{\xi^{(2)}(t; F^{(2)})\}$ , ... represent independent, purely random processes such that the infinite convolution

$$(84) \quad F^{(1)} * F^{(2)} * \dots$$

is convergent. Then the sum  $\{\xi^{(1)}(t; F^{(1)})\} + \{\xi^{(2)}(t; F^{(2)})\} + \dots$  will be convergent, and constitute a purely random process with the convolution (84) for principal distribution function.

In fact, the convergent convolution (84) is the distribution function of the sum  $\sum_{i=1}^{\infty} \xi^{(i)}(t; F^{(i)})$ , which is thus convergent. According to a remark attached to theorem 1, the convergence of this sum implies the convergence of  $\sum_{i=1}^{\infty} \{\xi^{(i)}(t; F^{(i)})\}$ .

A characteristic property of the purely random process is that the two variables  $\xi_C(t_1, \dots, t_n)$  and  $\xi(t_1, \dots, t_n)$  will have identical distribution functions for any condition ( $C$ ) not referring to any time point in the set  $(t) = (t_1, \dots, t_n)$ . Thus we have in this case (cf. (57))

$$E_C[\xi(t_{m+1})] = E[\xi(t)].$$

Any random series, e.g. a series of records on throws with a die, will form a model series of the purely random process. The illustrative model series used in the present study are very simply constructed, the double purpose being to facilitate the calculations, and to bring into relief the characteristic features of the different types of process. For this construction, the well-known random sampling numbers of L. H. C. TIPPETT (1927) were used.

Two independent random series, denoted by  $(\bar{a}_t^{(1)})$  and  $(\bar{a}_t^{(2)})$ , will be given for illustration. Denoting the corresponding processes by  $\{\alpha^{(1)}(t; F_1)\}$  and  $\{\alpha^{(2)}(t; F_2)\}$  respectively, the defining function  $F_1$  is given by

$$\begin{aligned} F_1(u) &= 0 \text{ for } u < -1, & F_1(u) &= 1 \text{ for } -1 \leq u < 0, \\ F_1(u) &= .9 \text{ for } 0 \leq u < 1, & F_1(u) &= 1 \text{ for } 1 \leq u, \end{aligned}$$

and the function  $F_2$  by

$$\begin{aligned} F_2(u) &= 0 \text{ for } u < -1, & F_2(u) &= .3 \text{ for } -1 \leq u < 0, \\ F_2(u) &= .7 \text{ for } 0 \leq u < 1, & F_2(u) &= 1 \text{ for } 1 \leq u. \end{aligned}$$

A short calculation shows that the mean value of each process  $\{\alpha_t^{(i)}\}$  equals zero, and that the variances, say  $D_1^2$  and  $D_2^2$ , are

$$(85) \quad D_1^2 = .2, \quad D_2^2 = .6.$$

Writing  $\tau$  for the TIPPETT numbers, the model series  $(\bar{\alpha}_t^{(1)})$  consists of the 1000 elements obtained from the first 1000  $\tau$ -numbers on page 1 by the use of the following code:

$$\bar{\alpha}_t^{(1)} = 1 \text{ for } \tau = 0; \quad \bar{\alpha}_t^{(1)} = 0 \text{ for } \tau = 1, \dots, 8; \quad \bar{\alpha}_t^{(1)} = -1 \text{ for } \tau = 9.$$

The second series was obtained from the corresponding  $\tau$ -numbers on page 2. The code used was

$$\bar{\alpha}_t^{(2)} = 1 \text{ for } \tau = 0, 1, 2; \quad \bar{\alpha}_t^{(2)} = 0 \text{ for } \tau = 3, \dots, 6;$$

$$\bar{\alpha}_t^{(2)} = -1 \text{ for } \tau = 7, 8, 9.$$

The first 100 elements in each of the  $\bar{\alpha}$ -series will now be quoted.

Table 1. (1) Model series  $(\bar{\alpha}_t^{(1)})$ ; first 100 elements.

0	-1	0	0	0	0	0	0	-1	-1	0	-1	0	-1	0	-1	0	-1
0	-1	0	0	0	0	0	1	0	0	0	0	0	0	0	-1	0	0
0	0	0	0	0	0	-1	0	0	0	1	0	0	0	0	0	0	1
0	0	-1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	1	0	-1	0	-1	0	0	0	0	-1	0	1	0

(2) Model series  $(\bar{\alpha}_t^{(2)})$ ; first 100 elements.

1	1	0	0	1	-1	0	-1	-1	0	0	-1	0	1	1	0	0	-1	0
-1	0	-1	0	1	0	1	-1	0	0	-1	-1	0	0	0	0	-1	1	1
1	-1	1	1	-1	1	0	0	0	-1	-1	-1	1	0	1	1	-1	-1	1
1	1	1	1	1	0	1	0	1	1	0	1	1	1	1	-1	-1	-1	-1
0	-1	-1	0	0	0	1	1	1	-1	0	0	1	0	-1	0	-1	1	-1

The  $\bar{\alpha}$ -series will be used repeatedly in the present study. For this reason each of the series, considered a statistical population with hypothetical distribution functions  $F_1(u)$  and  $F_2(u)$  respectively, has been tested as to the goodness of fit between the empirical and the theoretical distributions. Denoting the empirical distribution functions by  $\bar{F}_1(u)$  and  $\bar{F}_2(u)$  respectively, these were found to be

$$(86) \quad \begin{cases} \bar{F}_1(u) = 0 \text{ for } u < -1, \\ \bar{F}_1(u) = .908 \text{ for } 0 \leq u < 1, \end{cases} \quad \begin{cases} \bar{F}_1(u) = 100 \text{ for } -1 \leq u < 0, \\ \bar{F}_1(u) = 1 \text{ for } 1 \leq u; \end{cases}$$

$$(87) \quad \begin{cases} \bar{F}_2(u) = 0 \text{ for } u < -1, \\ \bar{F}_2(u) = .692 \text{ for } 0 < u < 1, \end{cases} \quad \begin{cases} \bar{F}_2(u) = 321 \text{ for } -1 \leq u < 0, \\ \bar{F}_2(u) = 1 \text{ for } 1 \leq u. \end{cases}$$

The  $\omega^2$ -test indicates a nice fit. As is readily verified by the insertion of (86) and (87) in formulae (322) and (323), the two  $\bar{\alpha}$ -series give the  $\omega^2$ -values '000064 and '000505, while the corresponding expectations are '000180 and '000420.

The model series ( $\bar{\alpha}$ ) have also been tested with regard to the concordance between serial ( $\bar{r}_k$ ) and autocorrelation ( $r_k$ ) coefficients. The latter vanish for  $k \neq 0$ . On the other hand, writing  $n$  for the number of elements in the series, and paying no regard to terms of order  $1/n$ , the sampling dispersion of any autocorrelation coefficient of the two series is found to be  $1/\sqrt{n} = .032$ . The first five serial coefficients are given below.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\bar{r}_k(\bar{\alpha}_t^{(1)})$	'057	'047	'010	'015	'036
$\bar{r}_k(\bar{\alpha}_t^{(2)})$	'046	'011	'006	'004	'004

The deviation from the corresponding autocorrelation coefficient is in no case larger than the double dispersion. It is rather interesting to note that although the series consist of as many as 1000 elements, a serial coefficient amounting to '06 cannot be considered significantly positive.

The first 20 values of the series  $(\bar{\alpha}_t^{(1)} - \bar{\alpha}_t^{(2)})$  obtained from table 1 are given below to illustrate a linear operation with independent random processes:

$$(88) \quad -1 -2 0 0 -1 1 0 1 1 -1 -1 1 -1 -1 -2 0 0 -1 1 -1$$

The series thus obtained is a model series of the process  $\{\alpha^{(1)} - \alpha^{(2)}\}$ . Since in the present case  $\{\alpha^{(1)} + \alpha^{(2)}\}$  is found to possess the same defining distribution function as  $\{\alpha^{(1)} - \alpha^{(2)}\}$ , the series (88) also forms a model series for the process  $\{\alpha^{(1)} + \alpha^{(2)}\}$ .

We shall next pass to some other type cases of the discrete stationary process, denoted by  $\beta$ ,  $\gamma$  and  $\delta$ , which will be built up by successive linear operations on the purely random processes.

$\beta.$  *The process of moving averages.* According to the analysis in section 13, a stationary process  $\{\xi(t)\}$  will be obtained by taking

$$(89) \quad \xi(t) = b_0 \cdot \eta(t) + b_1 \cdot \eta(t-1) + \cdots + b_h \cdot \eta(t-h),$$

letting  $\{\eta(t)\}$  represent a purely random process, say  $\{\eta(t; F_\eta(u))\}$ , and  $(b)=(b_0, b_1, \dots, b_h)$  an arbitrary sequence of real numbers. The type of process thus defined will be called *the process of moving averages*. A specific process (89) of this type will sometimes be denoted by  $\{\xi(t; \eta)\}$ . The purely random process  $\{\eta(t)\}$  and the variables  $\eta(t_1, \dots, t_n)$  will be called *primary* in respect of  $\{\xi(t; \eta)\}$  and  $\xi(t_1, \dots, t_n)$  respectively.

Let it be observed that, for any constant  $c > 0$ , the variable (89) is identical with that defined by the variables  $\eta(t; F_\eta(c \cdot u))$  and by the sequence  $c \cdot (b)=(c \cdot b_0, c \cdot b_1, \dots, c \cdot b_h)$ . Therefore, the assumption  $b_0=1$  often imposed on (89) in the following will not restrict the generality of the analysis. On the other hand, if  $\sum b_i \neq 0$ , we can find an identical process such that  $\sum b_i = 1$ . Hence the name proposed for the process.

The principal distribution functions  $F_\xi(u)$  and  $F_\eta(u)$  of  $\{\xi(t)\}$  and  $\{\eta(t)\}$  respectively are connected by the relation

$$(90) \quad F_\xi(u) = F_\eta(u/b_0) * F_\eta(u/b_1) * \cdots * F_\eta(u/b_h).$$

In case  $D(\eta(t))$  is finite, we obtain further

$$D^2(\xi(t)) = (b_0^2 + b_1^2 + \cdots + b_h^2) \cdot D^2(\eta(t)).$$

For exemplifying conditioned variables connected with a process  $\{\xi(t; \eta)\}$  of moving averages, let  $m < h$ , and let  $C$  denote the condition

$$(C) = (\eta(t-m-1) = \eta_{m+1}, \eta(t-m-2) = \eta_{m+2}, \dots, \eta(t-h) = \eta_h).$$

Then the conditioned variable  $\xi_C(t)$  will be given by

$$\xi_C(t) = b_0 \cdot \eta(t) + \cdots + b_m \cdot \eta(t-m) + b_{m+1} \cdot \eta_{m+1} + \cdots + b_h \cdot \eta_h.$$

As is readily verified, we have in case  $D(\eta(t))$  is finite,

$$(91) \quad E_C[\xi(t)] = (b_0 + b_1 + \cdots + b_m) E[\eta(t)] + b_{m+1} \cdot \eta_{m+1} + \cdots + b_h \cdot \eta_h.$$

Model series for the process of moving averages are readily obtained by applying a moving linear operation to a model series for the purely random process. For instance, the series of differences  $\Delta^k \bar{\eta}_t$ , of any order  $k$ ; obtained from a purely random series  $\bar{\eta}_t$  will illustrate the process of moving averages. Below, as an illustrative model series are given the first 100 values of  $\Delta \bar{\alpha}_t^{(2)} = \bar{\alpha}_{t+1}^{(2)} - \bar{\alpha}_t^{(2)}$  obtained from Table 1. The corresponding process will be denoted

$$(92) \quad \{\beta(t; \alpha^{(2)})\} = \{\alpha^{(2)}(t+1; F_2) - \alpha^{(2)}(t; F_2)\}.$$

Table 2. Model series ( $\bar{\beta}_t$ ); first 100 elements.

0	-1	0	1	-2	1	-1	0	1	0	-1	1	1	0	-1	0	0	-1	1	-1
1	-1	1	1	-1	1	-2	1	0	-1	0	1	0	0	0	-1	2	0	0	
-2	2	0	-2	2	-1	0	0	-1	0	0	2	-1	1	0	-2	0	2	0	0
0	0	0	-1	1	-1	1	0	-1	1	0	-1	1	0	0	-2	0	0	0	1
-1	0	1	0	0	1	0	0	-2	1	0	1	-1	-1	1	-1	0	2	-2	2

γ. *The general process of linear regression.* Let  $\{\eta(t)\}$  stand for a purely random process with finite dispersion  $D(\eta)$ , and let  $b_0, b_1, b_2, \dots$  be a real sequence such that  $\sum_{k=0}^{\infty} b_k^2$  is convergent. Finally, we must assume either that  $E[\eta] = 0$  or that  $\sum_{k=0}^{\infty} b_k$  be convergent.

These alternatives will present themselves repeatedly in the sequel. The former assumption is better suited to our purpose; the modifications caused by the latter are trivial. Accordingly, we shall always assume that  $E[\eta] = 0$ .

Considering the series

$$(93) \quad b_0 \eta(t) + b_1 \eta(t-1) + b_2 \eta(t-2) + \dots,$$

it follows from the independence of the variables  $\eta(t)$  that the variance of

$$(94) \quad b_n \eta(t-n) + b_{n+1} \eta(t-n-1) + \dots + b_{n+p} \eta(t-n-p)$$

is given by

$$(b_n^2 + b_{n+1}^2 + \dots + b_{n+p}^2) \cdot D^2(\eta).$$

The dispersion of (94) thus tends to zero uniformly in  $p$  as  $n \rightarrow \infty$ . Accordingly, the sum (93) is convergent (cf. p. 41). As remarked in connexion with theorem 1, it follows that we may define a stationary process  $\{\xi(t)\}$  by writing

$$(95) \quad \{\xi(t)\} = b_0 \{\eta(t)\} + b_1 \{\eta(t-1)\} + b_2 \{\eta(t-2)\} + \dots$$

By definition, this is the general formula for a *process of linear regression*.

Having dealt under article 6 with a particular case of the general process of linear regression, we shall in the next article build up the scheme of autoregression as another type case. As a matter of fact, in the present study the general process will be chiefly used as a convenient tool for comprehension of the two type cases mentioned, and it is only these which will appear in the applications. It will also be sufficient to refer to the type cases for model series.

6. *The process of linear autoregression.* Let  $(a) = (a_1, \dots, a_h)$  stand for real numbers such that  $a_h \neq 0$ , and that the roots of the equation (34) all are of a modulus less than 1. Let further  $(b) = (b_1, b_2, \dots)$  be a sequence such that (A) the difference equation

$$(96) \quad x(t) + a_1 \cdot x(t-1) + \dots + a_h \cdot x(t-h) = 0$$

is satisfied when  $x(t) \equiv b_t$ , and (B) the initial values  $b_1, \dots, b_h$  are solutions of the following system of linear equations

$$(97) \quad \begin{cases} a_1 + b_1 = 0 \\ a_2 + a_1 \cdot b_1 + b_2 = 0 \\ \vdots \vdots \vdots \vdots \vdots \vdots \vdots \vdots \\ a_{h-1} + a_{h-2} \cdot b_1 + \dots + a_1 \cdot b_{h-2} + b_{h-1} = 0 \\ a_h + a_{h-1} \cdot b_1 + \dots + a_1 \cdot b_{h-1} + b_h = 0. \end{cases}$$

Since the determinant of the system (97) is given by

$$\begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ a_1 & 1 & 0 & \cdots & 0 & 0 \\ a_2 & a_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{h-1} & a_{h-2} & \cdots & \cdots & a_1 & 1 \end{vmatrix}$$

and thus equals 1, the initial  $b$ -values are uniquely determined. It is also seen that all  $b_i$  are real. Moreover, the series  $\sum_0^\infty b_s^2$  is convergent (cf. section 6). Letting  $\{\eta(t)\}$  represent a purely random process with finite dispersion, the conditions indicated under article  $\gamma$  thus will be satisfied. Hence, a stationary process  $\{\xi(t)\} = \{\xi(t; \eta)\}$  will be defined by putting

$$(98) \quad \xi(t; \eta) = \eta(t) + b_1 \cdot \eta(t-1) + b_2 \cdot \eta(t-2) + \dots$$

By definition, this operation gives rise to the general *process of (linear) autoregression*. Since  $a_h \neq 0$ , the autoregression will be said to be of *order h*.

As pointed out in section 13, the variables  $\zeta(t)$  defined by the following linear operation on the variables  $\xi(t; \eta)$  given by (98) will likewise constitute a stationary process:

$$\zeta(t) = \xi(t) + a_1 \cdot \xi(t-1) + \dots + a_h \cdot \xi(t-h).$$

It will now be shown that the process  $\{\zeta(t)\}$  thus defined is equivalent with  $\{\eta(t)\}$ . The proof is based on a transformation of a double sum of aleatory variables.

By definition, we have

$$\zeta(t) = \sum_{i=0}^h a_i \sum_{k=0}^\infty b_k \cdot \eta(t-i-k) = \lim_{N \rightarrow \infty} \sum_{i=0}^h a_i \sum_{k=0}^N b_k \cdot \eta(t-i-k),$$

where  $a_0$  and  $b_0$  should be given the value 1. Introducing an auxiliary variable  $\zeta_N(t)$ , an elementary transformation shows that

$$\begin{aligned} \zeta_N(t) &= \sum_{i=0}^h a_i \sum_{k=0}^N b_k \eta(t-i-k) = \sum_{p=0}^{h-1} \eta(t-p) \sum_{q=0}^p a_{p-q} b_q + \sum_{p=h}^N \eta(t-p) \sum_{q=0}^h a_q b_{p-q} + \\ &+ \sum_{p=N+1}^{N+h} \eta(t-p) \sum_{q=p-N}^h a_q b_{p-q} = \eta(t) + c_1 \cdot \eta(t-N-1) + \dots + c_h \cdot \eta(t-N-h), \end{aligned}$$

where the second transformation is a consequence of (96) and (97). Putting  $a = \max |a_i|$ , the coefficients  $c_s$  introduced are seen to satisfy the inequality

$$(99) \quad |c_s| \leq a \cdot (|b_{N+s-h}| + |b_{N+s+1-h}| + \dots + |b_{N-1}| + |b_N|).$$

Paying regard to the convergence of  $\sum b_i^2$ , we conclude without difficulty that  $c_1 \eta(t-N-1) + \dots + c_h \eta(t-N-h)$  tends to zero

in probability as  $N \rightarrow \infty$ . Thus,  $\zeta_N(t)$  tends to  $\eta(t)$  in probability. According to the corollary of theorem 1, the proof that  $\{\zeta(t)\}$  equals  $\{\eta(t)\}$  is thereby completed, and we get the following fundamental identity,

$$(100) \quad \{\xi(t)\} + a_1 \cdot \{\xi(t-1)\} + \cdots + a_h \cdot \{\xi(t-h)\} = \{\eta(t)\}.$$

The relation (100) implies, i. a., that

$$(101) \quad \xi(t) + a_1 \cdot \xi(t-1) + \cdots + a_h \cdot \xi(t-h) = \eta(t).$$

Observing further that  $\eta(t)$  is independent of  $\eta(t-1), \eta(t-2), \dots$ , and — according to (98) — also of  $\xi(t-1), \xi(t-2), \dots$ , the relation (101) shows that the variables  $\xi(t), \xi(t-1), \dots, \xi(t-h)$  are connected by a relation of linear regression. Hence the name proposed for the process.

A simple illustration of conditioned variables is given by

$$\xi_C(t) = \eta(t) - a_1 \xi_{t-1} - a_2 \xi_{t-2} - \cdots - a_h \xi_{t-h},$$

where  $(C) = (\xi(t-1) = \xi_{t-1}, \xi(t-2) = \xi_{t-2}, \dots, \xi(t-h) = \xi_{t-h})$ . More general formulae will be given in section 23.

Construction of model series for a process of autoregression, given for example by (98), may be performed in the same way as in the case of a finite moving average of a purely random series. The difficulty of an infinite number of weights  $b_s$  is but apparent, for when a certain precision in the calculations is fixed, only a finite number of the weights, say  $H$  of them, will be found to have any influence.

Denoting by  $\bar{d}_t$  the values in a model series of the present type, and representing the primary model series by  $(\bar{a}_t)$ , the formula for the construction reads (cf. (63))

$$\bar{d}_t = \bar{a}_t + b_1 \cdot \bar{a}_{t-1} + b_2 \cdot \bar{a}_{t-2} + \cdots + b_H \cdot \bar{a}_{t-H}.$$

Having constructed  $\bar{d}_1, \bar{d}_2, \dots, \bar{d}_H$  according to this formula, the subsequent values  $\bar{d}_{H+1}, \bar{d}_{H+2}$ , etc. may be obtained from the more convenient recurrence formula

$$(102) \quad \bar{d}_t = \bar{a}_t - a_1 \cdot \bar{d}_{t-1} - a_2 \cdot \bar{d}_{t-2} - \cdots - a_h \cdot \bar{d}_{t-h}.$$

In the illustrative series given below, a slight simplification has been made, in that formula (102) has been applied also for  $t=1$ ,

$2, \dots, H$ , taking  $\bar{\delta}_t = 0$  for  $t < 1$ . As the series consist of 1000 elements each, this modification of the first few elements will not have any disturbing effect upon serial coefficients and other quantities relating to the whole of the series. On the other hand, thanks to the modification adopted, the construction of the model series may be readily followed in detail.

Three illustrative series, denoted by  $\{\bar{\delta}_t^{(1)}\}$ ,  $\{\bar{\delta}_t^{(2)}\}$  and  $\{\bar{\delta}_t^{(3)}\}$  respectively, will be presented. The formulae of type (100) for the corresponding processes read

$$(103) \quad \{\delta^{(1)}(t)\} = \{\alpha^{(2)}(t)\} - .8 \{\delta^{(1)}(t-1)\},$$

$$(104) \quad \{\delta^{(2)}(t)\} = \{\alpha^{(1)}(t)\} + .8 \{\delta^{(2)}(t-1)\},$$

$$(105) \quad \{\delta^{(3)}(t)\} = \{\alpha^{(2)}(t)\} + .2 \{\delta^{(3)}(t-1)\} - .65 \{\delta^{(3)}(t-2)\}.$$

The verification of the recurrent calculation of the  $\bar{\delta}$ -series needs no comment.

Table 3. (1) Model series  $(\bar{\delta}_t^{(1)})$ . First 50 elements.

1.00	.20	- .16	.13	.90	-1.72	1.37	-2.10	.68	- .54
.44	-1.35	1.08	.14	.90	- .72	.57	- .46	- .64	.51
-1.41	1.12	-1.90	1.52	- .22	.17	.86	-1.69	1.35	-1.08
- .13	- .89	.71	- .57	.46	- .37	.29	-1.23	1.99	- .59
1.47	-2.18	2.74	-1.19	- .04	1.04	- .83	.66	- .53	- .58

(2) Model series  $(\bar{\delta}_t^{(2)})$ . First 50 elements.

.00	-1.00	- .80	- .64	- .51	- .41	- .33	- .26	- .21	-1.17
-1.93	-1.55	-2.24	-1.79	-2.43	-1.95	-1.56	-2.25	-1.80	-2.44
-1.95	-2.56	-2.05	-1.64	-1.31	-1.05	- .84	- .33	.26	.21
.17	.14	.11	.09	.07	.06	- .95	- .76	- .61	- .49
- .39	- .31	- .25	- .20	- .16	- .13	-1.10	- .88	- .71	- .56

(3) Model series  $(\bar{\delta}_t^{(3)})$ . First 50 elements.

1.00	1.20	- .41	- .86	1.09	- .22	- .76	-1.01	- .71	.51
.56	-1.22	- .61	1.67	1.73	- .74	-1.27	.23	- .13	- .17
- .95	- .08	- .40	- .03	1.25	.27	.24	-1.13	- .38	.66
- .62	-1.55	.09	1.03	.15	- .64	- .22	- .63	1.02	1.61
.66	-1.92	.19	2.28	- .67	- .62	- .31	.47	- .11	-1.32

$\pi$ . *On the periodic processes.* In this article, a stationary process will be constructed which belongs to the class of singular processes as introduced in section 14. A distribution function  $F(u)$  and an integer  $h$  being arbitrarily given, there will be constructed a stationary process, say  $\{\xi(t)\}$ , with the following properties: (a) the process will be singular by means of the relation

$$(106) \quad \xi(t) - \xi(t-h) = 0,$$

(b) the process has  $F(u)$  for principal distribution function. According to the construction, any sample series, say  $(\xi_t, \xi_{t+1}, \dots)$ , will be strictly periodic, and with period  $h$ . The process will, accordingly, be termed a *periodic process*.

The simple construction device reads as follows. Let  $\xi = [\xi^{(1)}, \dots, \xi^{(h)}]$  represent an  $h$ -dimensional aleatory variable such that (A) the distribution function of  $\xi$ , say  $F(u_1, \dots, u_n)$ , is symmetrical in respect of the variables  $u_i$ , (B) all the variables  $\xi^{(i)}$  have  $F(u)$  for distribution function. Now, let a sequence of multi-dimensional aleatory variables be defined by

It is evident that these variables may be taken for the variables

$$\xi(t), \quad \xi(t, t-1), \dots, \xi(t, t-1, \dots, t-h+1),$$

$$\xi(t, \dots, t-h+1, t-h), \xi(t, \dots, t-h+1, t-h, t-h-1), \dots,$$

$$\xi(t, \dots, t-h+1, t-h, \dots, t-2h+1),$$

connected with a stationary process  $\{\xi(t)\}$ . It needs no comment that this process has the advanced properties (a) and (b).

For exemplifying conditioned probability distributions connected with the periodic process constructed above, let  $C$  stand for a condition implying  $\xi(t) = \xi_t$ , and let  $n$  represent an integer. Then  $P_C[\xi(t + n \cdot h) \in \theta] = 1$  if  $\xi_t$  belongs to the set  $\theta$ , otherwise  $P_C = 0$ . Further,  $E_C[\xi(t + n \cdot h)] = \xi_t$ .

For the construction of a model series, say  $\pi_1, \pi_2, \dots$ , illustrating the periodic process defined above, it will be sufficient to form a

model random sample  $\bar{\pi}_1, \bar{\pi}_2, \dots, \bar{\pi}_h$  on the basis of the distribution function  $F(u)$ , and then take  $\bar{\pi}_{h+1} = \bar{\pi}_1, \bar{\pi}_{h+2} = \bar{\pi}_2$ , etc. Still simpler, a model series of this type has been obtained by letting a coin-throw decide whether  $\bar{\pi}_1$  should equal 1 or -1, and by taking  $\bar{\pi}_{2i} = -\bar{\pi}_1, \bar{\pi}_{2i+1} = \bar{\pi}_1$ . The resulting series reads

$$(107) \quad 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ -1 \ 1 \ \dots$$

This series will be referred to as the  $\bar{\pi}$ -series.

Considering the difference equation (14) corresponding to the relation of singularity (106) of the periodic process constructed, this equation has for general solution a sum of harmonics, viz. the expression (15), a circumstance in full agreement with theorem 2. This aspect corresponds to interpreting a sample series  $\pi_t$  by means of FOURIER analysis as a composed harmonic (cf. section 5).

Having so far shown that the class of singular stationary processes introduced in section 14 is not vacuous, the next question is if there exist singular processes which satisfy relations (77) of a more general type than (106). As a matter of fact, any particular relation generalizing (106) will impose certain conditions upon the distribution functions  $\{F\}$  of the singular process. Leaving open the question of which special distribution functions may present themselves in case of a special singularity of type (77), we advance that the normal process (see section 16) will be found to admit any relation (83) as a singular case. After this reference to a process of superposed harmonics, only one of the schemes surveyed in Chapter I, viz. the scheme of hidden periodicities, remains to be interpreted as a stationary process.

*Q. The process of hidden periodicities.* Let  $\{\xi^{(1)}(t)\}, \dots, \{\xi^{(k)}(t)\}$  represent independent stationary processes. According to section 13, the sum  $\{\xi(t)\} = \{\xi^{(1)}(t)\} + \dots + \{\xi^{(k)}(t)\}$  will constitute a stationary process. If at least one of the processes  $\{\xi^{(i)}(t)\}$  is a periodic process, or a process of superposed harmonics,  $\{\xi(t)\}$  will be called a *process of hidden periodicities*. In particular, letting  $k=2$ , and taking for  $\{\xi^{(1)}(t)\}$  a process of superposed harmonics, and for  $\{\xi^{(2)}(t)\}$  a purely random process, we get the simple scheme of hidden periodicities dealt with in section 8.

A model series for the process of hidden periodicities may be obtained from independently constructed model series for the purely random process and the periodic process. Taking one model series

of each type, the series obtained by summing corresponding elements will form a model series for the process of hidden periodicities. The table below gives the first elements in two model series for the processes  $\{\Omega^{(1)}\}$  and  $\{\Omega^{(2)}\}$  defined by

$$\Omega^{(i)}(t) = \pi(t) + \alpha^{(i)}(t),$$

the  $\alpha$ - and  $\pi$ -processes being defined in corresponding articles of the present section. The two  $\bar{\Omega}$ -series consist of 1000 elements each. The construction may be followed in detail by means of table 1 and the series (107) (cf. formula (39)).

Table 4. (1) Model series  $(\bar{\Omega}_t^{(1)})$ ; first 100 elements.

1	-2	1	-1	1	-1	1	-1	1	-2	0	-1	0	-1	0	-1	1	-2	1	-2
1	-2	1	-1	1	-1	1	0	1	-1	1	-1	1	-1	1	-1	0	-1	1	-1
1	-1	1	-1	1	-1	0	-1	1	-1	2	-1	1	-1	1	-1	1	-1	2	0
1	-1	0	-1	1	-1	1	0	1	-1	1	-1	1	-1	2	-1	1	-1	1	-1
1	-1	1	-1	1	0	1	-2	1	-2	1	-1	1	-1	0	-1	2	-1	1	0

(2) Model series  $(\bar{\Omega}_t^{(2)})$ ; first 100 elements.

2	0	1	-1	2	-2	1	-2	0	-1	1	-2	1	0	2	-1	1	-1	0	-1
0	-1	0	-1	2	-1	2	-2	1	-1	0	-2	1	-1	1	-1	1	-2	2	0
2	-2	2	0	0	0	1	-1	1	-2	0	-2	2	-1	2	0	0	-2	2	0
2	0	2	0	1	0	1	0	2	-1	2	0	1	0	2	0	0	-2	0	-2
1	-2	0	-1	1	-1	2	0	2	-2	1	-1	2	-1	0	-1	0	-2	2	-2

In spite of the fact that the process of hidden periodicities is not the only type of sum of independent processes dealt with in the sequel, it seems superfluous to go into further details in this introductory section. In conclusion, the following simple relations involving conditioned variables and expectations will be mentioned

$$\Omega_C(t + n \cdot h) = \pi_t + \xi(t + n \cdot h).$$

$$E_C[\Omega(t + nh)] = \pi_t + E[\xi].$$

Here  $\{\pi(t)\}$  and  $\{\xi(t)\}$  represent independent stationary processes with sum  $\{\Omega(t)\}$ . The process  $\{\pi(t)\}$  is assumed to be periodic with period  $h$ , while  $n$  is an arbitrary integer, and  $C$  stands for the condition  $(C) = (\pi(t) = \pi_t)$ .

### 16. On the normal stationary process.

With reference to A. KOLMOGOROFF, A. KHINTCHINE in a paper (1934) recurred to in the next section touches upon a continuous stationary process constructed by means of normal distribution functions. *Mutatis mutandis*, the same construction device will supply a discrete stationary process. In the sequel, this process will be termed the *normal process*. As a basis for illustrating the general stationary process, the normal process will prove very useful. In fact, in spite of the formal developments connected with the general normal process being of a simple structure, the normal process will, by proper specializations, be able to illustrate any type of stationary process mentioned in the previous section, and besides — as already advanced — the singular processes satisfying relations of type (83).

Before going into details concerning the normal stationary process, it will be convenient to introduce the concept of a general normal distribution in an enumerable set of variables.

Let an infinite quadratic form with real coefficients be given by

$$(108) \quad Q(X_1, X_2, \dots) = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} \mu_{pq} \cdot X_p X_q.$$

In the following analysis, the variables  $X_i$  may take on any real values. Under such circumstances, the form  $Q$  will not always be convergent. If divergent, the form must be interpreted symbolically, viz. as the comprehension of all finite forms of type

$$Q_n(X_1, \dots, X_n) = Q(X_1, X_2, \dots, X_n, 0, 0, \dots).$$

Thus prepared, let (A)  $\mu_{pq} = \mu_{qp}$ , and (B) any determinant  $\Delta(n)$  of type

$$\Delta(n) = \begin{vmatrix} \mu_{11}, & \mu_{12}, & \dots, & \mu_{1n} \\ \mu_{21}, & \mu_{22}, & \dots, & \mu_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1}, & \mu_{n2}, & \dots, & \mu_{nn} \end{vmatrix}$$

be non-negative. Then, taking for  $(m) = (m_1, m_2, \dots)$  an arbitrary real sequence, a function  $f(X_1, X_2, \dots)$  in an enumerable set of variables  $(X) = (X_1, X_2, \dots)$  will be defined by

$$(109) \quad f(X_1, X_2, \dots) = e^{i \sum_{p=1}^{\infty} m_p \cdot X_p - \frac{1}{2} \cdot Q(X_1, X_2, \dots)}$$

As often as  $Q$  is divergent, this function must be interpreted symbolically in the same way as  $Q$ . Now, the function  $f_n(X_1, \dots, X_n)$  defined by

$$f_n(X_1, \dots, X_n) = f(X_1, \dots, X_n, 0, 0, \dots) = e^{i \sum_{p=1}^n m_p \cdot X_p - \frac{1}{2} \cdot Q_n(X_1, \dots, X_n)}$$

is the characteristic function of a certain normal,  $n$ -dimensional aleatory variable, say  $[\xi^{(1)}, \dots, \xi^{(n)}]$ , (see e. g. H. CRAMÉR (1937) p. 109). We have  $E[\xi^{(i)}] = m_i$ , and  $E[(\xi^{(i)} - m_i)(\xi^{(k)} - m_k)] = \mu_{ik}$ . The form  $Q_n$  will thus be of the type (76).

In case  $\mathcal{A}(n) \neq 0$ , the variable  $[\xi^{(1)}, \dots, \xi^{(n)}]$  will possess an absolutely continuous distribution, and the density function, say  $\varphi_n(u_1, \dots, u_n)$ , will be given by

$$\varphi_n(u_1, \dots, u_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\mathcal{A}(n)}} \cdot e^{-\frac{1}{2} q_n(u_1, \dots, u_n)}.$$

Here  $q_n(u_1, \dots, u_n)$  is defined by

$$q_n(u_1, \dots, u_n) = \sum_{p=1}^n \sum_{q=1}^n \frac{\mathcal{A}_{pq}(n)}{\mathcal{A}(n)} \cdot (u_p - m_p)(u_q - m_q),$$

where  $\mathcal{A}_{pq}(n)$  stands for the minor of  $\mathcal{A}(n)$  belonging to the element  $m_{pq}$ . Considering, on the other hand, the case  $\mathcal{A}(n) = 0$ , let it be assumed that  $\mathcal{A}(n)$  is of rank  $h < n$ . Writing the relations of singularity on the form (69) after a suitable arrangement of the variables  $\xi^{(i)}$ , it follows from the previous analysis that the variable  $[\zeta^{(1)}, \dots, \zeta^{(n)}]$  defined by (72) will be composed of  $h$ , and only  $h$ , non-constant variables  $\zeta^{(i)}$ . On the other hand, for  $i > h$  the variables  $\zeta^{(i)}$  reduce to the constants  $m_i^*$  given by (75). Expressing the characteristic function of  $[\zeta^{(1)}, \dots, \zeta^{(n)}]$  in the variables  $Z_i$ , this will reduce to

$$(110) \quad f_n^*(Z_1, \dots, Z_h) = e^{i \sum_{p=1}^h m_p^* \cdot Z_p - \frac{1}{2} Q_h^*(Z_1, \dots, Z_h)}$$

Next, let us consider the infinite sequence  $\xi = [\xi^{(1)}, \xi^{(2)}, \dots]$ . By construction, every finite sub-group  $[\xi^{(i_1)}, \dots, \xi^{(i_k)}]$  will possess a well-

defined probability distribution, and, further, satisfy all consistency relations of type (53—54). Thus  $\xi$  will constitute a random variable in an infinite number of dimensions. A variable  $\xi$  of this type will be termed *normal*. The function  $f$  will be called the *characteristic function* of  $\xi$ . This term is justified by the evident fact that the distribution of  $\xi$  is uniquely determined by  $f$ , and *vice versa*.

A necessary and sufficient condition that a normal distribution as defined by (109) may be taken to define a variable  $\xi(t, t-1, \dots)$  connected with a stationary process  $\{\xi(t)\}$  is that (a)  $m_p$  reduces to a constant, say  $m$ , and (b)  $\mu_{p,q}$  is a function of  $p - q$ , say  $\mu_{p-q}$ . In fact, taking  $X_t, X_{t-1}, X_{t-2}, \dots$  for variables in the characteristic function (109), the coefficients of  $X_{t-p}$  and of  $X_{t-p} \cdot X_{t-q}$  will be independent of  $t$  when, and only when, the conditions (a) and (b) are satisfied simultaneously.

According to the condition (A) attached to (108), we have  $\mu_{p-q} = \mu_{q-p} = \mu_{|p-q|}$ . Further, disregarding the case of empty determinants  $\mathcal{A}(n)$ , the second condition,  $\mathcal{A}(n) \geq 0$ , is seen to imply  $\mu_0 > 0$ . Thus a set of real numbers  $r_k = r_{-k}$  will be well-defined by putting

$$r_k = \mu_k / \mu_0.$$

In terms of these  $r_k$ , the conditions  $\mathcal{A}(n) \geq 0$  will reduce to the inequalities (81).

According to the above, the general formula for the characteristic function of the variable  $\xi(t, t-1, \dots) = \xi[t]$  connected with a normal stationary process  $\{\xi(t)\}$  is given by

$$(111) \quad f(X_t, X_{t-1}, \dots) = e^{im \cdot \sum_{p=0}^{\infty} X_{t-p} - \frac{\sigma^2}{2} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} r_{|p-q|} \cdot X_{t-p} X_{t-q}},$$

where (A)  $m, \sigma > 0$ , and  $r_k$  are real, and (B) the coefficients  $r_k$  satisfy the inequalities (81), viz.  $\mathcal{A}(r, n) \geq 0$ .

It is seen that the normal stationary process defined by (111) has for principal distribution function the normal distribution function

$$\Phi(u) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{1}{2}\sigma^2(x-m)^2} dx,$$

and that the coefficients  $r_k$  appearing in (111) are nothing else than the autocorrelation coefficients of the process.

Thanks to theorem 2, the existence of a singular normal process satisfying (83) will be proved if the autocorrelation coefficients of any composed harmonic satisfying (30) are such that

$$(112) \quad \sum_{p=0}^n \sum_{q=0}^n r_{|p-q|} \cdot X_{t-p} X_{t-q} \geq 0$$

for any  $n$ , and for any real sequence  $(X) = (X_t, X_{t+1}, X_{t-2}, \dots)$ . For then, this form, and the quantity  $m$  obtained from (30), satisfy the conditions for introduction in the general formula (111) of the characteristic function of the variables  $\xi(t, t-1, \dots)$  defining the normal process, while the dispersion  $\sigma > 0$  may be chosen freely.

The remaining proof involves no difficulty. Denoting by  $x(t)$  an arbitrary composed harmonic satisfying (30), let a set of integers be given by  $0 \leq i_1 < i_2 < \dots < i_n$ . Employing an argument used by A. KHINTCHINE (1934), the obvious relations

$$(113) \quad 0 \leq \frac{1}{N} \cdot \sum_{s=1}^N [X_{t-i_1}(x(t-i_1+s)-m) + X_{t-i_2}(x(t-i_2+s)-m) + \dots + X_{t-i_n}(x(t-i_n+s)-m)]^2 = \\ = \sum_{p=1}^n \sum_{q=1}^n X_{t-i_p} \cdot X_{t-i_q} \cdot \frac{1}{N} \sum_{s=1}^N (x(t-i_p+s)-m)(x(t-i_q+s)-m)$$

will define a non-negative definite quadratic form in the variables  $X_{t-i_p}$ . After this observation, let  $N \rightarrow \infty$ . Then, disregarding a constant factor, the coefficient of  $X_{t-i_p} \cdot X_{t-i_q}$  in the quadratic form will, by definition, tend to the autocorrelation coefficient  $r_{|i_p-i_q|}$  which belongs to the composed harmonic considered (cf. (12)). Since the form will remain  $\geq 0$  also in the limit, and since the sequence  $i_p$  is arbitrary, the limit inequality implies (112).

Some circumstances connected with the singular processes merit particular attention. In the first place, assuming  $L[\xi(t) - m] = 0$  as given by (77) to be the relation of singularity of the lowest possible order, the distribution of  $\xi(t-1, \dots, t-h)$  will be non-singular. In other words, if in a sample series

$$\xi_i(t_0+k, t_0+k-1, \dots, t_0-h) = (\xi_{t_0+k}, \xi_{t_0+k-1}, \dots, \xi_{t_0-h})$$

the values  $\xi_{t_0-h}, \dots, \xi_{t_0-1}$  are known, neither  $\xi_{t_0-1}$  nor  $\xi_{t_0}, \xi_{t_0+1}, \dots$  will be uniquely determined. On the other hand, if the sample series section  $\xi_{t_0-h}, \dots, \xi_{t_0-1}$  is known, then  $\xi_{t_0}$ , and, recurrently

$\xi_{t_0+1}, \dots, \xi_{t_0+h}$  will, with probability one, be uniquely determined by means of  $L[\xi_{t_0+t} - m] = 0$ . Thus,  $\xi_{t_0+t}$  may be regarded as a solution of the difference equation (32) subject to the initial conditions  $\xi_{t_0-1}, \dots, \xi_{t_0-h}$ . According to the theory of difference equations (cf. section 6), the periods  $p_k$  of the individual harmonics in  $\xi_{t_0+t}$  will be determined by the coefficients  $a_i$  in (77), and therefore be the same in all sample series connected with the process considered. On the other hand, the amplitudes  $C_k$  and the phases  $\varphi_k$  will be determined by the initial values. As mentioned above, these contain a random element, so the amplitudes and the phases of the individual harmonics constituting  $\xi_{t_0+t}$  will vary from one sample series ( $\dots, \xi_{t_0+k}, \xi_{t_0+k-1}, \dots$ ) to another. Of course, any expectation connected with the varying phases and amplitudes, e.g.  $E[C_k^2]$ , forms a characteristic of the process, and may, considering e.g. a normal process defined by (111), be expressed in terms of  $m$ ,  $\sigma$  and  $r_k$  (cf. p. 73).

The above remarks apply, of course, to every singular process, and thus both to the periodic processes and to the singular normal processes. If the distribution of  $\xi(t-1, \dots, t-h)$  is absolutely continuous, which is always the case in the normal process, a more precise conclusion may be arrived at. In fact, let it in such case be assumed that not all of the individual harmonics in (17) would be present in  $\xi_{t_0+t}$ , i.e. that at least one harmonic would have a vanishing amplitude  $C_k$ . Then  $\xi_{t_0+t}$  must satisfy a difference equation of order  $h-1$  having a general solution satisfying also (32). Since there are only  $h$  such equations at most, the sample series sections  $(\xi_{t_0-1}, \dots, \xi_{t_0-h})$  having the property assumed will form a set of BOREL measure zero in the space of  $\xi(t-1, \dots, t-h)$ . Keeping in mind the absolute continuity assumed, it follows that with probability one all individual harmonics really will present themselves when writing a sample series  $\xi_{t_0+t}$  on the form (17).

The investigations of E. SLUTSKY (see (1927) and, e.g., (1937)) and V. ROMANOVSKY ((1932), (1933)) concerning the »sinusoidal limit law« fall under the theory of the stationary process, and present some parallelism with the previous analysis of the concept of singular process as introduced in section 14. Translating into the terminology of the present study, these authors investigate certain sequences of stationary processes, say  $\{\beta^{(1)}(t)\}, \{\beta^{(2)}(t)\}, \dots$ . Denoting by  $r_t^{(p)}$  the autocorrelation coefficients of the process  $\{\beta^{(p)}(t)\}$ , and by  $x(t)$  a function (17) satisfying a linear relation  $L[x(t) - m] = 0$ ,

the conditions imposed on the sequence  $\{\beta^{(p)}(t)\}$  imply that  $L[r_t^{(p)}] \rightarrow 0$  as  $p \rightarrow \infty$  (see V. ROMANOVSKY (1932), Theorem D). Representing by  $(\beta_1^{(p)}, \dots, \beta_n^{(p)})$  a section in a sample series of  $\{\beta^{(p)}(t)\}$ , and holding  $n$  fixed, the sinusoidal limit theorem asserts that, for sufficiently large values of  $p$ , the section considered will, with a probability as close to one as desired, approximate a function of type  $x(t)$  with any prescribed accuracy.

A few reflections on the previous analysis will verify this theorem. Writing  $m_p = E[\beta^{(p)}(t)]$ , let an auxiliary set of processes  $\{\xi^{(1)}(t)\}, \{\xi^{(2)}(t)\}, \dots$  be defined by  $\xi^{(p)}(t) = L[\beta^{(p)}(t) - m_p] = A^{2s} \beta^{(p)}(t-s) + \dots + h_s \cdot [\beta^{(p)}(t) - m_p]$ . It follows from the conditions imposed on  $r_t^{(p)}$  that the variables  $\xi^{(p)}(t+k)$  will tend in probability to zero as  $p \rightarrow \infty$ . It remains to prove that the composite variable  $[\xi^{(p)}(t), \xi^{(p)}(t+1), \dots, \xi^{(p)}(t+n)]$  will, for any fixed  $n$ , tend in probability to  $(0, 0, \dots, 0)$  as  $p \rightarrow \infty$ . This, however, follows at once from the Boole inequality (64).

By examples of the type  $\{\beta^{(p)}(t)\} = \sum_{i=1}^p a_{pi} \cdot \{\alpha(t-i)\}$ , E. SLUTSKY and V. ROMANOVSKY show that their theorems are not empty. While the added processes  $\alpha$  used by SLUTSKY are all of the purely random type, ROMANOVSKY gives other examples as well. The recent paper of E. SLUTSKY (1937) already referred to contains references to certain related investigations, and illustrates in full detail the behaviour of model series of the processes  $\{\beta^{(p)}(t)\}$  considered. In full agreement with the sinusoidal limit theorem, sections of moderate length in a model series approximate composed harmonics (17) of the proper type. The periods of the individual harmonics are the same in different sections, while the amplitudes and phases vary. These features are seen to be analogous to the properties of the singular processes proved above in connexion with the singular normal process. This parallelism is not accidental. In fact, an analysis of the sequences studied by E. SLUTSKY and V. ROMANOVSKY will show that these are convergent, and that the limit processes are singular. — In section 25 is given, i. a., a general device for the construction of sequences  $\{\beta^{(p)}(t)\}$  ruled by the sinusoidal limit theorem.

### 17. The autocorrelation coefficients as FOURIER constants.

In a paper already referred to, A. KHINTCHINE (1934) studies what in a continuous statistical process with finite dispersion corresponds to the autocorrelation coefficients in a discrete process, viz. the function defined for any real  $u$  by

$$R(u) = E[(\xi(t) - E[\xi]) \cdot (\xi(t+u) - E[\xi])] / D^2(\xi),$$

where  $\{\xi(t)\}$  represents the continuous stationary process considered. Terming  $R(u)$  the *correlation function* of the process, KHINTCHINE gives, *i. a.*, a necessary and sufficient condition that a function  $R(u)$  be the correlation function of a continuous stationary process. Slightly modifying the result, the condition is that there exists a distribution function, say  $V(x)$ , such that  $V(0) = 0$ , and

$$(114) \quad R(u) = \int_0^\infty \cos ux \cdot dV(x).$$

The inversion formula (see e. g. H. CRAMÉR (1935), Theorem 9)

$$V(x) = \frac{2}{\pi} \int_0^\infty R(u) \frac{\sin ux}{u} du$$

shows that the function  $V(x)$  is uniquely determined by  $R(u)$ .

We shall first give a corresponding theorem on the discrete stationary process. It should be observed that the same theorem holds for the generalized process (cf. section 11).

*Theorem 5.* Let  $r_k$  ( $k = 0, \pm 1, \pm 2, \dots$ ) be an arbitrary sequence of constants. A necessary and sufficient condition that there exists a discrete stationary process with the  $r_k$ 's for autocorrelation coefficients is that the  $r_k$ 's are the FOURIER constants of a non-decreasing function, say  $W(x)$ , such that  $W(0) = 0$ ;  $W(\pi) = \pi$ ,

$$(115) \quad r_k = \frac{1}{\pi} \int_0^\pi \cos kx \cdot dW(x).$$

Before passing to the proof we give the following inversion formula involving a converging sum (see e. g. F. HAUSDORFF (1923) p. 245)

$$(116) \quad W(x) = x + 2 \cdot \sum_{k=1}^{\infty} \frac{r_k}{k} \sin kx.$$

With suitable agreements as to points of discontinuity, the function  $W(x)$  thus will be uniquely determined by the autocorrelation coefficients. In the sequel,  $W(x)$  will be called the *generating function* of the autocorrelation coefficients  $r_k$ .

The proof will use some facts concerning definite quadratic forms, facts parallel to the properties of definite functions used by A. KHINTCHINE in the continuous case. In other respects the proofs are coincident.

For a verification of the necessity of the condition given, let  $\{\xi(t)\}$  be an arbitrary discrete stationary process with finite dispersion, say  $\sigma$ . Put  $E[\xi(t)] = m$ , let  $r_k$  represent the autocorrelation coefficients of the process  $\{\xi(t)\}$ , and consider the quadratic form

$$(117) \quad \sum_{p=0}^n \sum_{q=0}^n r_{|p-q|} \cdot X_p X_q, \quad n = 0, 1, 2, \dots$$

For any real sequence  $X_{t_1}, X_{t_2}, \dots, X_{t_n}$  we have (cf. (113))

$$\begin{aligned} 0 &\leq \frac{1}{\sigma^2} E \left[ \sum_{p=1}^n X_{t_p} [\xi(t_p) - m] \right]^2 = \\ &= \frac{1}{\sigma^2} \cdot \sum_{p=1}^n \sum_{q=1}^n E [X_{t_p} X_{t_q} [\xi(t_p) - m] [\xi(t_q) - m]] = \sum_{p=1}^n \sum_{q=1}^n r_{|t_p - t_q|} \cdot X_{t_p} X_{t_q}. \end{aligned}$$

This relation implies that the forms (117) are non-negative definite. Next, according to a theorem of G. HERGLOTZ<sup>1</sup> (1911), this statement is equivalent to saying that the following system of trigonometrical moments,

$$(118) \quad \frac{1}{2\pi} \int_0^{2\pi} \cos kx \cdot dW(x) = r_k; \quad \frac{1}{2\pi} \int_0^{2\pi} \sin kx \cdot dW(x) = 0,$$

has a non-decreasing solution  $W(x)$  with  $W(0) = 0$ .

The inversion of (118) gives exactly (116). The required relation (115) follows directly from (116), and since (116) further gives

---

<sup>1</sup> Prof. T. CARLEMAN has kindly informed me that the formulae (118) may be obtained directly from the HILBERT representation of a definite quadratic form of general type.

$$W(2\pi - x) = 2\pi - x - 2 \cdot \sum_{k=1}^{\infty} \frac{r_k}{k} \sin kx,$$

we obtain  $W(x) = 2\pi - W(2\pi - x)$ , and, finally,  $W(\pi) = \pi$ .

On the other hand, if  $r_k$  is given by (115), the HERGLOTZ theorem asserts that the forms (117) are non-negative. Thus, the relation (112) holds, and there exists a normal process with the given  $r_k$ -values for autocorrelation coefficients.

The theorem proved above permits of some general conclusions concerning the autocorrelation coefficients of a discrete stationary process. Considering the coefficients  $r_k$  for large  $k$ -values, their behaviour will depend on the continuity structure of the generating function  $W(x)$ . In order to study the situation in some detail, let

$$W(x) = a^{(1)} \cdot W^{(1)}(x) + a^{(2)} \cdot W^{(2)}(x) + a^{(3)} \cdot W^{(3)}(x)$$

stand for the well-known (cf. e.g. H. CRAMÉR (1928), p. 59) representation of  $W(x)$  as the sum of three uniquely determined, non-decreasing functions such that  $W^{(1)}(0) = W^{(2)}(0) = W^{(3)}(0)$ ,  $W^{(1)}(\pi) = W^{(2)}(\pi) = W^{(3)}(\pi) = \pi$ ,  $a^{(1)} + a^{(2)} + a^{(3)} = 1$ ,  $a^{(i)} \geq 0$ , and

$$1) \quad a^{(1)} \cdot W^{(1)}(x) = \int_0^\pi \frac{dW(y)}{dy} \cdot dy,$$

2)  $a^{(2)} \cdot W^{(2)}(x)$ , the *saltus* function, is equal to the sum of saltuses of  $W(x)$  at all the points of discontinuity which are less than or equal to  $x$ ; writing  $\lambda_v$  for the saltus points of  $W(x)$ , and  $c_v^2 \cdot \pi/2$  for the corresponding saltuses, then

$$\frac{a^{(2)}}{\pi} \cdot W^{(2)}(x) = \sum_{\lambda_v \leq x} c_v^2 / 2,$$

3)  $a^{(3)} \cdot W^{(3)}(x)$ , the *singular* function, is a continuous function which has almost everywhere a derivative equal to zero.

Thus prepared, we put

$$r_k^{(i)} = \frac{1}{\pi} \int_0^\pi \cos kx \cdot dW^{(i)}(x), \quad i = 1, 2, 3,$$

and obtain

$$(119) \quad r_k = a^{(1)} \cdot r_k^{(1)} + a^{(2)} \cdot r_k^{(2)} + a^{(3)} \cdot r_k^{(3)}.$$

The components  $r_k^{(i)}$  thus uniquely determined by the  $r_k$ -sequence

via its generating function (cf. (116)) are of entirely different character.

As to  $r_k^{(1)}$  we have (see e. g. H. C. Carslaw (1930), p. 271)

$$a^{(1)} \cdot r_k^{(1)} = \frac{1}{\pi} \int_0^\pi \frac{dW(x)}{dx} \cos kx dx \rightarrow 0 \text{ as } k \rightarrow \infty.$$

The component  $r_k^{(2)}$  is given by

$$(120) \quad a^{(2)} \cdot r_k^{(2)} = \frac{1}{2} \cdot \sum_{i=1}^{\infty} c_i^2 \cdot \cos \lambda_i k.$$

It is seen that  $r_k^{(2)}$  is an almost periodic function. Again referring to (19), we conclude that arbitrarily large  $k$ -values exist for which  $r_k^{(2)}$  approximates

$$r_0^{(2)} = \frac{1}{\pi} W^{(2)}(\pi) = \frac{1}{2 a^{(2)}} \cdot \sum_{i=1}^{\infty} c_i^2.$$

The singular component  $r_k^{(s)}$  permits of no unconditioned statement as to its behaviour for large  $k$ -values.

In order to arrive at a criterion of the structure of  $W(x)$ , let it be assumed that  $\sum_{k=1}^{\infty} |r_k|$  is convergent. It follows from (116) that in such a case  $W(x)$  for all  $x$  has a derivative  $W'(x)$ , that this will be obtained by summing the derivatives of the terms in the right member of (116), and that  $W'(x)$  will be bounded. We thus obtain the following corollary to theorem 5.

*Corollary.* Let  $\{\xi(t)\}$  be a stationary process with autocorrelation coefficients  $r_k$  such that  $\sum |r_k|$  is convergent. Then  $W(x)$  will be absolutely continuous,

$$W(x) = W^{(1)}(x).$$

The derivative  $W'(x)$  is bounded in modulus, and given by

$$(121) \quad W'(x) = \sum_{k=-\infty}^{\infty} r_k \cos kx, \quad 0 \leq x < \pi.$$

As a first application of the above analysis, we shall touch upon some problems concerning the relation between continuous and discrete stationary processes. The question to be put corresponds to a problem dealt with by G. Elfving (1937) in a study on MARKOFF chains.

A continuous stationary process, say  $\{\xi^{(c)}(t)\}$ , gives a hypothetical scheme for the probability relations in any time points by means of a set  $\{F^{(c)}\}$  of distribution functions  $F(t_1, \dots, t_n; u_1, \dots, u_n)$  satisfying (53)–(55) and referring to quite arbitrary time points  $t_1, \dots, t_n$ . Among these distribution functions, let those referring to integral time points be represented by  $\{F^{(d)}\}$ . It is plain that the set  $\{F^{(d)}\}$  thus obtained will define a discrete stationary process, say  $\{\xi^{(d)}(t)\}$ . The situation may be described by saying that the hypothesis  $\{\xi^{(c)}(t)\}$  is consistent with the hypothesis  $\{\xi^{(d)}(t)\}$ .

Marking the symbols referring to consistent processes  $\{\xi^{(c)}\}$  and  $\{\xi^{(d)}\}$  by (c) and (d) respectively, let  $\{\xi^{(c)}\}$  have a finite dispersion. Then, evidently, we have for any integral  $k$  (cf. (114))

$$r_k^{(d)} = R^{(c)}(k).$$

Further, for any  $x$  in the interval  $(0, \pi)$ , we have

$$(122) \quad W(x) = \pi \cdot \sum_{n=0}^{\infty} [V(n \cdot 2\pi + x) - V(n \cdot 2\pi - x)].$$

In fact, inserting the right member of (122) in (115), and paying regard to (114), we obtain by elementary transformations for integral  $k$ -values

$$r_k^{(d)} = \int_0^\pi \cos kx \cdot \sum_{n=0}^{\infty} d[V(n \cdot 2\pi + x) - V(n \cdot 2\pi - x)] = R^{(c)}(k).$$

On the other hand, let a discrete process  $\{\xi^{(d)}(t)\}$  be defined by a set  $\{F^{(d)}\}$  of distribution functions referring to integral time points. Our question is whether there exists an »interpolating» continuous process  $\{\xi^{(c)}(t)\}$  defined by a set  $\{F^{(c)}\}$  where the distribution functions referring to integral time points are identical with the given set  $\{F^{(d)}\}$ . Studying in the first place the autocorrelation coefficients  $r_k^{(d)} = \frac{1}{\pi} \int_0^\pi \cos kx dW(x)$ , we seek a continuous stationary process  $\{\xi^{(c)}\}$  with correlation function  $R^{(c)}(u)$  such that, for integral  $k$ -values,  $R^{(c)}(k) = r_k^{(d)}$ . Taking  $V(x) = \frac{1}{\pi} W(x)$  for  $0 \leq x \leq \pi$ , and  $V(x) = 1$  for  $x \geq \pi$ , such a function is evidently yielded by

$$(123) \quad R^{(c)}(u) = \int_0^\infty \cos ux \cdot dV(x).$$

It should be observed that (123) remains unchanged for integral  $u$ -values when substituting, for instance,  $V(x - n \cdot 2\pi)$  for  $V(x)$ , letting  $n$  denote a positive integer. Thus, the interpolating function  $R^{(c)}(u)$  will by no means be uniquely determined by the autocorrelation coefficients  $r_k^{(d)}$  given. This indeterminateness is, of course, analogous to the circumstance mentioned in section 4 that there exists an infinite number of simple harmonics all of which pass through all of the values  $x(t_n)$  taken on in equidistant points  $t_n$  by a simple harmonic  $x(t)$ .

In case the process  $\{\xi^{(d)}\}$  considered is normal, the KHINTCHINE-KOLMOGOROFF device for constructing a normal continuous process may be applied on the basis of an interpolating  $R^{(c)}(u)$  as given, for instance, by (123). Furnishing the normal continuous process with the same mean and the same dispersion as  $\{\xi^{(d)}\}$ , the resulting process will possess the property desired, i.e. give rise to the primary discrete process  $\{\xi^{(d)}\}$  when considering the probability relations in integral time points.

We conclude from the above that any theorem on the continuous stationary process also holds, *mutatis mutandis*, for the discrete process. On the other hand, since there possibly does not exist any continuous process interpolating an arbitrarily given discrete process, a theorem concerning the latter may be vacuous and trivial when applied to continuous processes.

*Illustration.* Taking  $V(x) = 1 - e^{-x}$ , formula (114) gives

$$R^{(c)}(u) = 1/(1 + u^2).$$

The sum (122) is readily computed, and gives

$$(124) \quad W(x) = \pi \left[ 1 - \frac{e^{-x} - e^{x-2\pi}}{1 - e^{-2\pi}} \right]; \quad 0 \leq x \leq \pi.$$

Insertion of this expression in (115) gives without difficulty

$$(125) \quad r_k^{(d)} = \frac{1}{1 + k^2} + \frac{2k \sin \pi k}{(e^\pi - e^{-\pi})(1 + k^2)}.$$

In full agreement with the general results, we find that for integral  $k$ -values  $r_k^{(d)} = R^{(c)}(k)$ .

On the other hand, in case a discrete process is given, and has  $W(x)$  as defined by (124) for generating function, one of the interpolating correlation functions will be obtained by taking  $R^{(c)}(u) = r_u^{(d)}$ , where  $r_k^{(d)}$  for  $-\infty < k < \infty$  is given by (125).

The corresponding function  $V(x)$  is, of course, equal to  $\frac{1}{\pi} W(x)$  in  $(0, \pi)$  and equal to 1 in  $(\pi, \infty)$ .

The second application of formula (115) is concerned with the periodogram method for graduating a sample series section connected with a stationary process. Only discrete processes will be considered; the arguments also hold, however, in the continuous case.

First, an investigation will be made as to whether a SCHUSTER periodogram analysis of a sample series section, say  $\xi_{t-1}, \dots, \xi_{t-n}$ , may be expected to be effective, in particular for large  $n$ -values. It will turn out that if the generating function of the process has a non-vanishing saltus component, the periodogram analysis will give positive results, viz. in the sense that the expectance of certain well-defined periodogram ordinates will be positive. On the other hand, in case the corollary of theorem 5 applies, a periodogram analysis will prove resultless.

Let  $(\xi_{t-1}, \xi_{t-2}, \dots)$  stand for a sample series belonging to a discrete stationary process  $\{\xi(t)\}$  with dispersion  $\sigma$ , mean  $m$ , and generating function  $W(x)$ .

Applying the classical SCHUSTER periodogram analysis to the sample series section  $(\xi_{t-1}, \xi_{t-2}, \dots, \xi_{t-n})$ , let the resulting periodogram functions be denoted by (see (26) and (27))

$$(126) \quad A(n, \lambda) = \frac{2}{n} \sum_{p=1}^n (\xi_{t_0+p} - m) \cos \lambda p; \quad B(n, \lambda) = \frac{2}{n} \sum_{q=1}^n (\xi_{t_0+q} - m) \sin \lambda q;$$

$$C^2(n, \lambda) = A^2(n, \lambda) + B^2(n, \lambda),$$

where  $t_0 = t - n - 1$ . Studying in the first place the expression

$$E = \lim_{n \rightarrow \infty} E[C^2(n, \lambda)],$$

elementary transformations yield

$$(127) \quad E[C^2(n, \lambda)] =$$

$$\begin{aligned} &= \frac{4}{n^2} \sum_{p=1}^n \sum_{q=1}^n (\cos \lambda p \cdot \cos \lambda q + \sin \lambda p \cdot \sin \lambda q) E[(\xi_{t_0+p} - m)(\xi_{t_0+q} - m)] = \\ &= \frac{4 \sigma^2}{n^2} \sum_{p=1}^n \sum_{q=1}^n r_{|p-q|} \cos \lambda(p-q) = \frac{4 \sigma^2}{n} \sum_{k=-n+1}^{n-1} \frac{n-|k|}{n} r_k \cos \lambda k = \\ &= \frac{4 \sigma^2}{n} \left[ 1 + 2 \sum_{k=1}^{n-1} \left( 1 - \frac{k}{n} \right) r_k \cos \lambda k \right]. \end{aligned}$$

In case the process  $\{\xi(t)\}$  is non-autocorrelated, we have  $r_k = 0$  for  $k > 0$ . The above formula then reduces to the SCHUSTER formula (37).

In a study on sampling problems in intercorrelated series, E. SLUTSKY (1934) investigates, i. a., expectations of type  $E[C^2(n, \lambda)]$  for  $\lambda$ -values equalling multiples of  $2\pi/n$ , i. e. expectations connected with the FOURIER coefficients of a sample series section  $(\xi_{t-1}, \dots, \xi_{t-n})$  (cf. (25)). Under certain restrictive conditions concerning the process considered, SLUTSKY gives the relations (127).

J. BARTELS (1935) seems to be the first to have deduced the relations (127) without setting restrictive conditions on the process analysed.

In order to avoid discussions which might obscure the point of the analysis, we shall now introduce a restriction concerning the generating function of the autocorrelation coefficients, viz. that  $a^{(3)} = 0$ , and that  $\Sigma |r_k^{(1)}|$  is convergent. According to (121), the latter assumption implies that  $dW^{(1)}(\lambda)/d\lambda$  is finite for all  $\lambda$ .

It is seen that the limit expectation  $E$  may be split up into portions corresponding to (119), say  $a^{(1)} \cdot E^{(1)}$ ,  $a^{(2)} \cdot E^{(2)}$  and  $a^{(3)} \cdot E^{(3)}$ . By the simplifying hypotheses made, we have  $a^{(3)} = 0$ .

Considering the relations (127), we observe that an elementary transformation gives

$$(128) \quad 1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) r_k \cos \lambda k = \frac{1}{n} \cdot \sum_{s=0}^{n-1} \sum_{k=-s}^s r_k \cos \lambda k,$$

and that the right member is a CESÀRO mean, viz. the arithmetical mean of the first  $n$  partial sums  $\sum_{-s}^s r_k \cos \lambda k$  of the series appearing in (121). After substitution of  $r_k^{(1)}$  for  $r_k$ , the expression (128) thus will tend to  $dW^{(1)}(\lambda)/d\lambda$  as  $n \rightarrow \infty$ . As by hypothesis this derivative is bounded in modulus, we conclude that for all  $\lambda$  in  $(0, \pi)$

$$(129) \quad E^{(1)} = \frac{4\sigma^2}{n} \cdot \frac{dW^{(1)}(\lambda)}{d\lambda} + O(1/n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Paying regard to (120), a similar argument gives

$$(130) \quad \begin{cases} E^{(2)} = \lim_{n \rightarrow \infty} \frac{4\sigma^2}{n} \left[ 1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) r_k^{(2)} \cos \lambda k \right] = 0 \text{ for } \lambda \neq \lambda_v, \\ E^{(2)} = c_v^2 \cdot \sigma^2 / a^{(2)} \text{ for } \lambda = \lambda_v < \pi. \end{cases}$$

Thus, while an ordinate  $C^2(n, \lambda_v)$  in the periodogram of a sample series may vary from one series to another, its expectation is by a simple limit relation connected with the saltus,  $\frac{\pi}{2}c_v^2$ , in  $\lambda = \lambda_v$  of the generating function  $W(\lambda)$  of the autocorrelation coefficients.

It is seen from (129)–(130) that, under the assumptions made, it is only if  $a^{(2)} > 0$  that a periodogram analysis of a sample series will be fruitful. Assuming that there are  $s$  discontinuities in  $W^{(2)}(\lambda)$ , the analysis will result in a composed harmonic,

$$x_n(t-k) = \sum_{\nu=1}^s A(n, \lambda_\nu) \cos(n+1-k)\lambda_\nu + \sum_{\nu=1}^s B(n, \lambda_\nu) \sin(n+1-k)\lambda_\nu,$$

approximating the section  $(\xi_{t-1}, \dots, \xi_{t-n})$  analysed, and with coefficients depending on the sample series considered and satisfying

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma^2} E[A^2(n, \lambda_\nu) + B^2(n, \lambda_\nu)] = c_\nu^2.$$

A standard measure of the deviation  $\xi_{t-k} - x_n(t-k)$  is given by the expectation  $E$  defined by

$$E = E \left[ \frac{1}{n} \cdot \sum_{k=1}^n [\xi(t-k) - x_n(t-k)]^2 \right].$$

Disregarding terms of order  $1/n$ , the coefficients  $A$  and  $B$  in  $x_n(t-k)$  will make  $E$  a minimum (cf. (43)),

$$E \cong (1 - \frac{1}{2} c_1^2 - \dots - \frac{1}{2} c_s^2) \cdot \sigma^2 = \left[ 1 - \frac{a^{(2)}}{\pi} \cdot W^{(2)}(\pi) \right] \cdot \sigma^2.$$

This relation shows clearly the scope of the periodogram analysis. In fact, in the special case  $W(x) \equiv W^{(2)}(x)$ , the approximating function  $x_n(t-k)$  will, for sufficiently large  $n$ -values, yield a fit as close as desired.

On the other hand, it follows from the previous analysis that in case  $\Sigma |r_k|$  is convergent, then  $W(x) \equiv W^{(1)}(x)$ , and  $dW(x)/dx$  is bounded. In this case, a periodogram analysis of the section  $(\xi_{t-1}, \xi_{t-2}, \dots)$  will be resultless (cf. (129)). Processes of this type have to be attacked by the use of entirely different methods. Now, a method at once suggesting itself is that of linear regression analysis. For instance, approximating  $\xi(t)$  linearly by means of  $\xi(t-1)$  we obtain  $\xi(t) = r_1 \cdot \xi(t-1) + \eta(t)$ , denoting by  $\eta(t)$  a residual variable with variance  $\sigma^2 \cdot (1 - r_1^2)$ . Otherwise expressed,  $r_1 \cdot \xi_{t-1}$  yields an effective prognosis of  $\xi_t$  with a squared dispersion equalling  $E[(\xi_t - r_1 \cdot \xi_{t-1})^2] = \sigma^2 \cdot (1 - r_1^2)$ . This simple instance is sufficient for exemplifying that this method is of quite another

type than the periodogram analysis. The next stage of the present study will be to follow up this line of research; this is done in section 19. The coming section is reserved for a preparatory survey.

For illustrations of periodogram analysis of model series, reference is given to section 25. For the present we only remark that the general formula (127) gives

$$(131) \quad \frac{1}{\pi} \int_0^\pi E[C^2(n; \lambda)] d\lambda = \frac{4\sigma^2}{n}.$$

Comparing with (37), the expectance of the ordinates in a periodogram is seen to be a function of  $\lambda$  with mean value equalling the constant (in respect of  $\lambda$ ) expectance in the case of a purely random process.

It is rather interesting to compare the relations (131) and (130). It is seen that (131) also holds in the case of a process of hidden periodicities, and that in such a process the ordinates  $E[C^2(n; \lambda_\nu)]$  will not tend to zero as  $n \rightarrow \infty$ . We conclude that the rise to a maximum in  $E[C^2(n; \lambda)]$  will be very rapid, and that the corresponding peaks in the periodogram will be very thin, with a breadth of an order of magnitude not surpassing  $1/n$ .

On the other hand, if the corollary to theorem 5 applies, the expectance  $E[C^2(n; \lambda)]$  evidently tends to zero uniformly in  $\lambda$  as  $n \rightarrow \infty$ ,

$$E[C^2(n; \lambda)] = \frac{4W'(\lambda)}{n} \cdot \sigma^2 + O(1/n).$$

In full agreement with (121) we conclude from (131) that in the present case  $\frac{1}{\pi} \int_0^\pi W'(\lambda) d\lambda = 1$ .

### 18. On linear approximation in a space of random variables.

The next section presenting a generalization of the regression analysis of ordinary random variables to the general discrete stationary process, the present section is reserved for an interpretation of the ordinary regression analysis as a linear approximation in a metrical space. The interpretation applies, of course, to statistical as well as to aleatory random variables.

When dealing in the following with a set of random variables, say  $\xi$ , we shall tacitly assume that any finite sub-group, say  $[\xi^{(1)}, \dots,$

$\xi^{(n)}$ ], forms a well-defined multi-dimensional variable, say with distribution function  $F(u_1, \dots, u_n)$ , and that the functions  $F$  are consistent (cf. p. 32 f.). As observed by M. FRÉCHET ((1937), p. 205 f.), a set of one-dimensional random variables with finite dispersion can be made metrical by defining the distance between two variables in the set, say  $\xi^{(i)}$  and  $\xi^{(k)}$ , as the dispersion of the difference variable  $\xi^{(i)} - \xi^{(k)}$ ,

$$(132) \quad |\xi^{(i)} - \xi^{(k)}| = D(\xi^{(i)} - \xi^{(k)}).$$

This fact depends on the simple inequality

$$(133) \quad D(\xi^{(1)} - \xi^{(2)}) + D(\xi^{(1)} - \xi^{(3)}) \geq D(\xi^{(2)} - \xi^{(3)}),$$

which by elementary transformations reduces to the inequality of SCHWARZ.

Adopting the distance definition (132), next let  $a$  stand for a real number, and let  $\xi$  and  $\eta$  be two random variables. Then we have

$$(134) \quad |\xi - a \cdot \eta|^2 = D^2(\xi) - 2a \cdot r(\xi, \eta) \cdot D(\xi) \cdot D(\eta) + a^2 D^2(\eta).$$

This squared distance will reach a minimum equalling

$$(135) \quad [1 - r^2(\xi, \eta)] \cdot D^2(\xi)$$

when  $a$  equals the regression coefficient of  $\xi$  on  $\eta$ , i. e. when

$$(136) \quad a = r(\xi, \eta) \cdot D(\xi) / D(\eta).$$

This regression coefficient is linear in respect to  $\xi$ . In fact,

$$(137) \quad \begin{aligned} r(\xi + \zeta, \eta) \cdot D(\xi + \zeta) &= \frac{E[(\xi - E[\xi] + \zeta - E[\zeta])(\eta - E[\eta])]}{D(\eta)} = \\ &= \frac{E[(\xi - E[\xi])(\eta - E[\eta])] + E[(\zeta - E[\zeta])(\eta - E[\eta])]}{D(\eta)} = \\ &= r(\xi, \eta) \cdot D(\xi) + r(\zeta, \eta) \cdot D(\zeta). \end{aligned}$$

Using these properties of regression coefficients, the multiple regression theory founded and developed by the English statistical school (see G. U. YULE and M. G. KENDALL (1937), p. 511, for references) can be interpreted as a particular branch of the theory of approximations in general linear spaces. In the general terminology, uncorrelated variables  $\xi^{(i)}$  and  $\xi^{(k)}$  should be called orthogonal elements in the space to which they belong. For later application we shall record some general approximation formulae in terms of determinants. The verification being in detail parallel to the GRAM-SCHMIDT orthogonalization of vectors in an infinite number of dimensions, reference is again made to G. KOWALEWSKI ((1909), § 175).

Let an  $n$ -dimensional variable  $[\xi^{(1)}, \dots, \xi^{(n)}]$  formed by variables  $\xi^{(i)}$  with finite dispersion be linearly non-singular (see p. 41 f.). Considering an arbitrary variable  $\xi^{(0)}$  with finite dispersion, and writing  $m_i = E[\xi^{(i)}]$ ;  $\mu_{ik} = E[(\xi^{(i)} - m_i)(\xi^{(k)} - m_k)]$ , there exists a well-defined sequence of coefficients  $a_{in}$  minimizing the dispersion of the variable defined for arbitrary  $a_{in}$ 's by

$$(138) \quad \xi^{(0)} - m_0 - a_{1n} \cdot (\xi^{(1)} - m_1) - \dots - a_{nn} \cdot (\xi^{(n)} - m_n).$$

Terming *residual*, and denoting by  $\eta^{(n)}$  the variable (138) formed by the minimizing coefficients  $a_{in}$ , we have  $E[\eta^{(n)}] = 0$ , and

$$(139) \quad \eta^{(n)} = \frac{\begin{vmatrix} \mu_{11}, \dots, \mu_{1n}, \xi^{(1)} - m_1 \\ \dots \dots \dots \dots \dots \\ \mu_{n1}, \dots, \mu_{nn}, \xi^{(n)} - m_n \\ \mu_{01}, \dots, \mu_{0n}, \xi^{(0)} - m_0 \end{vmatrix}}{\begin{vmatrix} \mu_{11}, \dots, \mu_{1n} \\ \dots \dots \dots \\ \mu_{n1}, \dots, \mu_{nn} \end{vmatrix}}, \quad D^2(\eta^{(n)}) = \frac{\begin{vmatrix} \mu_{11}, \dots, \mu_{1n}, \mu_{10} \\ \dots \dots \dots \\ \mu_{n1}, \dots, \mu_{nn}, \mu_{n0} \\ \mu_{01}, \dots, \mu_{0n}, \mu_{00} \end{vmatrix}}{\begin{vmatrix} \mu_{11}, \dots, \mu_{1n} \\ \dots \dots \dots \\ \mu_{n1}, \dots, \mu_{nn} \end{vmatrix}}.$$

A set of formulae equivalent to (139), and involving an auxiliary set of one-dimensional aleatory variables  $\zeta^{(i)}$ , is given by

$$(140) \quad \eta^{(n)} = \xi^{(0)} - m_0 - c_1 \cdot \zeta^{(1)} - c_2 \cdot \zeta^{(2)} - \dots - c_n \cdot \zeta^{(n)},$$

$$(141) \quad D^2(\eta^{(n)}) = D^2(\xi^{(0)}) - c_1^2 - c_2^2 - \dots - c_n^2.$$

For the auxiliary variables we have (cf. G. KOWALEWSKI (1909) p. 426 and T. LINDBLAD (1937)),

$$(142) \quad \zeta^{(1)} = \frac{\xi^{(1)} - m_1}{\sqrt{1 \cdot \mu_{11}}}, \quad \zeta^{(2)} = \frac{\begin{vmatrix} \mu_{11}, \xi^{(1)} - m_1 \\ \mu_{21}, \xi^{(2)} - m_2 \end{vmatrix}}{\sqrt{\mu_{11} \cdot \begin{vmatrix} \mu_{11}, \mu_{12} \\ \mu_{21}, \mu_{22} \end{vmatrix}}};$$

$$\zeta^{(3)} = \frac{\begin{vmatrix} \mu_{11}, \mu_{12}, \xi^{(1)} - m_1 \\ \mu_{21}, \mu_{22}, \xi^{(2)} - m_2 \\ \mu_{31}, \mu_{32}, \xi^{(3)} - m_3 \end{vmatrix}}{\sqrt{\begin{vmatrix} \mu_{11}, \mu_{12} \\ \mu_{21}, \mu_{22} \end{vmatrix} \cdot \begin{vmatrix} \mu_{11}, \mu_{12}, \mu_{13} \\ \mu_{21}, \mu_{22}, \mu_{23} \\ \mu_{31}, \mu_{32}, \mu_{33} \end{vmatrix}}};$$

These variables are *standardized*, i. e.

$$(143) \quad E[\zeta^{(i)}] = 0; \quad D[\zeta^{(i)}] = 1.$$

They are further mutually non-correlated,

$$(144) \quad r(\zeta^{(i)}, \zeta^{(k)}) = E[\zeta^{(i)} \cdot \zeta^{(k)}] = 0 \text{ for } i \neq k.$$

and it is thanks to this relation that the coefficients  $c_i$  in (140) are independent of  $n$ ,

$$(145) \quad c_i = r(\xi^{(0)}, \zeta^{(i)}) \cdot D(\xi^{(0)}) = E[(\xi^{(0)} - m_0) \cdot \zeta^{(i)}].$$

On the other hand, in case  $[\xi^{(1)}, \dots, \xi^{(n)}]$  is singular, say of rank  $h$ , and on account of the relations (69), the coefficients  $a_{i,n}$  in the residual variable  $\eta^{(n)}$  will not be uniquely determined. In fact, the addition of an arbitrary linear combination of the vanishing sums (69) will not change the variable (138). In full agreement herewith, the expressions (139) become indeterminate when  $[\xi^{(1)}, \dots, \xi^{(n)}]$  is taken to be singular.

Whether or not  $[\xi^{(1)}, \dots, \xi^{(n)}]$  be singular, the residual variable  $\eta^{(n)}$  will be uncorrelated with every  $\xi^{(i)}$ . In fact, formulae (134)–(136) imply that otherwise the variance  $D^2(\eta^{(n)})$  could be brought down by subtraction in  $\eta^{(n)}$  of the non-vanishing variable

$$r(\eta^{(n)}, \xi^{(i)}) \frac{D(\eta^{(n)})}{D(\xi^{(i)})} \cdot [\xi^{(i)} - m_i].$$

Considering the residuals  $\eta^{(1)}, \eta^{(2)}, \dots, \eta^{(n)}$ , it follows from (141) that the variances  $D^2(\eta^{(1)}), D^2(\eta^{(2)}), \dots, D^2(\eta^{(n)})$  form a non-increasing sequence. It is of central importance that, when  $D^2(\eta^{(i)}) > 0$ , we shall have  $D^2(\eta^{(i+1)}) = D^2(\eta^{(i)})$  if, and only if, one of the following two cases is present:

(A)  $[\xi^{(1)}, \dots, \xi^{(i+1)}]$  is singular by means of a relation of type

$$(146) \quad p_1 \cdot (\xi^{(1)} - m_1) + \dots + p_i \cdot (\xi^{(i)} - m_i) = \xi^{(i+1)} - m_{i+1}$$

(B) No relation of type (146) exists, but  $\xi^{(i+1)}$  is uncorrelated with  $\eta^{(i)}$ .

It follows, i. a., that if  $[\xi^{(1)}, \dots, \xi^{(n)}]$  is singular of rank  $h$ , the variables  $\xi^{(i)}$  may be arranged in an order such that the coefficients  $a_{p,q}$  in the residuals  $\eta^{(1)}, \eta^{(2)}, \dots, \eta^{(h)}$ , and only in these, are uniquely determined, while

$$(147) \quad D(\xi^{(0)}) \geq D(\eta^{(1)}) \geq D(\eta^{(2)}) \geq \cdots \geq D(\eta^{(h)}) = \\ = D(\eta^{(h+1)}) = \cdots = D(\eta^{(n)}) \geq 0.$$

In other words, the first  $h$  variables  $\xi^{(i)}$  alone will be able to bring down the residual dispersion to its minimum value.

The following determinant expression for a general partial correlation coefficient in the notation of G. U. YULE (see G. U. YULE and M. G. KENDALL (1937), p. 269) will attach the above system of formulae to the familiar theory of multiple correlation. The formula is valid in case none of the variables  $[\xi^{(0)}, \xi^{(2)}, \xi^{(3)}, \dots, \xi^{(n)}]$  and  $[\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(n)}]$  is singular.

$$r_{01, 23 \dots n} = \frac{\left| \begin{array}{c} \mu_{01}, \mu_{02}, \dots, \mu_{0n} \\ \mu_{21}, \mu_{22}, \dots, \mu_{2n} \\ \vdots \quad \vdots \quad \ddots \\ \mu_{n1}, \mu_{n2}, \dots, \mu_{nn} \end{array} \right|}{\sqrt{\left| \begin{array}{cc} \mu_{11}, \mu_{12}, \dots, \mu_{1n} & \mu_{00}, \mu_{02}, \dots, \mu_{0n} \\ \mu_{21}, \mu_{22}, \dots, \mu_{2n} & \mu_{20}, \mu_{22}, \dots, \mu_{2n} \\ \vdots \quad \vdots \quad \ddots & \vdots \quad \vdots \quad \ddots \\ \mu_{n1}, \mu_{n2}, \dots, \mu_{nn} & \mu_{n0}, \mu_{n2}, \dots, \mu_{nn} \end{array} \right|}}.$$

A proof from  $n$  to  $n + 1$  will verify

$$D^2(\eta^{(n)}) = \mu_{00} \cdot (1 - r_{01}^2)(1 - r_{02}^2) \cdots (1 - r_{0n}^2)_{12 \dots (n-1)},$$

a formula which shows, i. a., that also the partial correlation coefficients lie in the interval  $(-1, 1)$ .

For later application, we record that

$$(148) \quad a_{1n} = r_{01, 23 \dots n} \cdot \frac{D_{0, 23 \dots n}}{D_{1, 23 \dots n}},$$

and analogous formulae hold for the remaining coefficients  $a_{in}$  (cf. G. U. YULE and M. G. KENDALL (1937), p. 266). Here  $D_{i, 23 \dots n}$  for  $i = 0$  or 1 represents the dispersion of the residual variable obtained when approximating  $\xi^{(i)}$  by the variables  $\xi^{(2)}, \xi^{(3)}, \dots, \xi^{(n)}$ .

For application in forecast problems, let us interpret (138) in terms of conditioned expectations. Writing  $(C) = (\xi^{(1)} = \xi_1, \dots, \xi^{(n)} = \xi_n)$ , an estimate of  $\xi_C^{(0)}$ , which is the best linear one in the sense of principle of the least squares, is yielded by

$$(149) \quad F_C[\xi^{(0)}] = m_0 + a_{1n}(\xi_1 - m_1) + \cdots + a_{nn}(\xi_n - m_n).$$

We have

$$\eta_C^{(0)} = \xi_C^{(0)} - F_C[\xi^{(0)}]$$

and

$$E[\xi_C^{(0)} - F_C[\xi^{(0)}]]^2 = E[\eta_C^{(0)}]^2 = D^2(\eta_C^{(0)}),$$

where the latter relation results from (57).

If the variables  $\xi^{(1)}, \dots, \xi^{(n)}$  are mutually uncorrelated, the coefficients  $a_{ik}$  are independent of  $k$ . Writing  $a_{ik} = a_i$ , and taking  $(C) = (\xi^{(i_1)} = \xi_{i_1}, \dots, \xi^{(i_k)} = \xi_{i_k})$ , we get in this case

$$(150) \quad F_C[\xi^{(0)}] = m_0 + a_{i_1}(\xi_{i_1} - m_{i_1}) + \cdots + a_{i_k}(\xi_{i_k} - m_{i_k}).$$

Further, if  $\eta^{(n)}$  is independent of the variables  $\xi^{(1)}, \dots, \xi^{(n)}$  we have

$$(151) \quad E_C[\eta^{(n)}] = E[\eta^{(n)}] = 0,$$

and

$$(152) \quad E_C[\xi^{(0)}] = m_0 + a_{1n}(\xi_1 - m_1) + \cdots + a_{nn}(\xi_n - m_n) + \\ + E_C[\eta^{(n)}] = F_C[\xi^{(0)}].$$

## 19. Linear autoregression analysis of the discrete stationary process.

In this section, the linear regression analysis as surveyed in the previous section will be applied to the variables  $\xi(t)$  connected with a stationary process  $\{\xi(t)\}$ . Approximating  $\xi(t)$  by means of  $\xi(t-1), \dots, \xi(t-n)$ , a well-defined procedure of consecutive approximations will be given. After a passage to the limit, we shall arrive at a residual variable with properties corresponding to the case of a finite number of approximations.

Let  $\{\xi(t)\}$  be a stationary process with finite dispersion  $\sigma$ , with mean  $m$ , and with principal correlation determinants  $A(r, n)$  given by (81). Approximating  $\xi(t)$  by means of  $\xi(t-1, \dots, t-n) = [\xi(t-1), \dots, \xi(t-n)]$  as described in the previous section, let the residual variable be given by

$$(153) \quad \eta(t; n) = \xi(t) - m - \alpha(1, n) \cdot [\xi(t-1) - m] - \alpha(2, n) \cdot [\xi(t-2) - m] - \cdots - \alpha(n, n) \cdot [\xi(t-n) - m].$$

In case  $\Delta(r, n-1) \neq 0$ , formula (139) yields

$$(154) \quad D^2(\xi(t)) \geq D^2(\eta(t; n)) = \sigma^2 \cdot \Delta(r, n) / \Delta(r, n-1) \geq 0.$$

Since  $D^2(\eta(t; n))$ ,  $n = 1, 2, \dots$ , forms a non-increasing sequence (cf. (147)), we get

$$(155) \quad 1 \geq \frac{\Delta(r, 1)}{1} \geq \frac{\Delta(r, 2)}{\Delta(r, 1)} \geq \cdots \geq \frac{\Delta(r, n)}{\Delta(r, n-1)} \geq \cdots \geq 0.$$

According to the analysis in section 14, one of two cases is present. Either  $\Delta(r, n)$  is above zero for all  $n$ , or  $\Delta(r, n) > 0$  for  $n < h$  while  $\Delta(r, n) = 0$  for  $n \geq h$ . The number  $h$  appearing in the latter case equals the rank of linear singularity of the process considered. Paying regard to these facts, and to the relation (155), it is evident that any stationary process belongs to one, and only to one, of the following classes:

(I). The process is non-singular, and there exists a positive constant  $x^2 \leq 1$  such that

$$(156) \quad D^2(\eta(t; n)) / \sigma^2 = \Delta(r, n) / \Delta(r, n-1) \rightarrow x^2 \leq 1 \text{ as } n \rightarrow \infty.$$

(II). The process is singular, say of rank  $h$ .

(III). The process presents no singularity of finite rank, but

$$(157) \quad D^2(\eta(t; n)) / \sigma^2 = \Delta(r, n) / \Delta(r, n-1) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

In this case, the process will be termed *singular of infinite rank*.

In the following analysis, we shall assume that the process considered belongs to the first class.

*Autoregression analysis of a non-singular process.* According to (140), the variables (153) may be written on the form

$$(158) \quad \eta(t; n) = \xi(t) - m - c_1 \xi_1^*(t) - c_2 \xi_2^*(t) - \cdots - c_n \xi_n^*(t),$$

where the standardized variables  $\xi_k^*(t)$  are given by (cf. (142))

$$(159) \quad \xi_n^*(t) = \frac{\begin{vmatrix} 1 & r_1 & \dots & r_{n-2} & \xi(t-1) - m \\ r_1 & 1 & \dots & r_{n-3} & \xi(t-2) - m \\ \dots & \dots & \dots & \dots & \dots \\ r_{n-1} & r_{n-2} & \dots & 1 & \xi(t-n) - m \end{vmatrix}}{\sigma \cdot V \Delta(r, n-1) \cdot \Delta(r, n-2)},$$

and where the coefficients  $c_i$  are uniquely determined, and independent of  $n$  (cf. (145)),

$$(160) \quad c_i = r[\xi(t), \xi^*(t)] \cdot \sigma.$$

According to the general analysis in section 13, the variables  $\xi_k^*(t)$  will, for an arbitrarily fixed  $k$ , constitute a stationary process  $\{\xi_k^*(t)\}$ . It is seen that the variable  $\xi_k^*[t]$  defined by this process will be of type (62). Similarly, for any  $n$  the variables  $\eta(t; n)$  will constitute a stationary process  $\{\eta(t; n)\}$  of the same type (62). It will next be shown that the sequence  $\{\eta(t; n)\}$  is convergent in probability as  $n \rightarrow \infty$ .

According to (141) and (156) we have in the first place

$$D^2(\eta(t; n)) = \sigma^2 - c_1^2 - \cdots - c_n^2 \rightarrow \kappa^2 \sigma^2 \text{ as } n \rightarrow \infty.$$

We conclude that  $\sum c_n^2$  is convergent, and that

$$(161) \quad c_1^2 + c_2^2 + \cdots = (1 - \kappa^2) \sigma^2.$$

Next, the variables  $\xi_1^*(t), \xi_2^*(t), \dots$  being uncorrelated (cf. (144)), we get

$$D^2(\eta(t; n+p) - \eta(t; n)) = c_{n+1}^2 + \cdots + c_{n+p}^2.$$

Keeping in mind that  $\sum c_i^2$  is convergent, it follows that this dispersion tends to zero uniformly in  $p$  as  $n \rightarrow \infty$ . Hence, paying regard to a remark in section 13 (see p. 40 f.), we conclude that the sequence  $\eta(t; 1), \eta(t; 2), \dots$  is convergent in probability. According to theorem 1, this implies that the sequence

$$(162) \quad \{\eta(t; 1)\}, \{\eta(t; 2)\}, \dots$$

will also converge in probability.

The limit process of the sequence (162) will be denoted  $\{\eta(t)\}$ . In analogy to the case of a finite number of approximations, the process  $\{\eta(t)\}$  and the corresponding variables  $\eta(t_1, \dots, t_n)$  will be termed *residual*. According to a remark on p. 41, the mean and dispersion of the residual process will be given by the corresponding limit characteristics of the sequence (162).

Observing that  $\eta(t; n)$  is non-correlated with the variables  $\xi(t-1), \dots, \xi(t-n)$ , and keeping in mind that the dispersion of  $\eta(t; n)$  lies above a positive constant, it follows readily that the limit residual  $\eta(t)$  is non-correlated with  $\xi(t-n)$  for any positive

integer  $n$ . Hence we conclude that  $\eta(t)$  for any  $k \geq 0$  is uncorrelated with all of the variables  $\xi_1^*(t-k)$ ,  $\xi_2^*(t-k)$ , ... defined by (159) (cf. (137)). Further we have

$$\begin{aligned} E[(\xi(t) - m) \cdot \eta(t)] &= \lim_{n \rightarrow \infty} E[(\xi(t) - m - c_1 \xi_1^*(t) - \cdots - c_n \xi_n^*(t)) \cdot \eta(t)] = \\ &= \lim_{n \rightarrow \infty} E[\eta(t; n) \cdot \eta(t)] = E[\eta(t) \cdot \eta(t)] = D^2(\eta(t)). \end{aligned}$$

Hence, paying regard to the relations (cf. p. 77)

$$E[\eta(t)] = E[\eta(t; n)] = 0,$$

we get

$$r(\xi(t), \eta(t)) = D(\eta(t)) / D(\xi(t)) = \alpha.$$

Writing generally

$$(163) \quad r(\xi(t+n), \eta(t)) / \alpha = b_n,$$

we thus have  $b_0 = 1$ , and  $b_n = 0$  for  $n < 0$ .

The above-mentioned properties of the residuals correspond directly with the finite case dealt with in the previous section. There is also another important analogy. Considering in the finite case the residuals obtained when approximating  $\xi^{(0)}$  and  $\xi^{(p)}$  by  $[\xi^{(1)}, \dots, \xi^{(n)}]$  and  $[\xi^{(p+1)}, \dots, \xi^{(n)}]$  respectively, a short reflection shows that these residuals are non-correlated. In order to show, correspondingly, that the residuals  $\eta(t)$  and  $\eta(t-p)$  are non-correlated, let it be observed that

$$\begin{aligned} E[\eta(t) \cdot \eta(t-p)] &= \lim_{k \rightarrow \infty} E[\eta(t; k) \cdot [\xi(t-p) - m - c_1 \xi_1^*(t-p) - \cdots - \\ &\quad - c_{k-p+1} \xi_{k-p+1}^*(t-p) - \cdots - c_k \xi_k^*(t-p)]] . \end{aligned}$$

Denoting by  $\varrho_{k-p}$  the sum appearing in the second row, we have  $E[\varrho_{k-p}] = 0$ , and  $D^2(\varrho_{k-p}) \rightarrow 0$  for any  $p > 0$  as  $k \rightarrow \infty$ . Thus

$$\begin{aligned} r(\eta(t), \eta(t-p)) &= \lim_{k \rightarrow \infty} r(\eta(t; k), \xi(t-p) - m - c_1 \xi_1^*(t-p) - \cdots - \\ &\quad - c_{k-p} \xi_{k-p}^*(t-p)). \end{aligned}$$

According to previous remarks, the correlation coefficient in the right member equals zero for any  $p > 0$ . Observing that

$$(164) \quad r(\eta(t), \eta(t+p)) = r(\eta(t-p), \eta(t)) = r(\eta(t), \eta(t-p)) = 0$$

we conclude that the process  $\{\eta(t)\}$  is non-autocorrelated.

Summing up the results, we get the following theorem.

*Theorem 6.* A residual process  $\{\eta(t)\}$  obtained from a non-singular stationary process  $\{\xi(t)\}$  is stationary and non-autocorrelated. The variable  $\eta(t)$  is non-correlated with  $\xi(t-1), \xi(t-2), \dots$ , while

$$r(\xi(t), \eta(t)) = D(\eta(t)) / D(\xi(t)).$$

The arguments used in the proof of this theorem also apply in the remaining cases II and III. As the residual variables  $\eta(t)$  are here seen to be vanishing, their correlation properties will be indeterminate. Accordingly, these cases need no further comment.

Illustrations of the autoregression analysis of stationary processes will be given in sections 25 and 26 (cf. also p. 92).

## 20. A canonical form of the discrete stationary process.

In this section it will be shown that the residual processes arrived at in the preceding section give a basis for the construction of a canonical form of the stationary process with finite dispersion.

Until further notice, the random variables considered will be assumed to have a vanishing mean.

In the case of a finite number of approximations dealt with in section 18, the following representation holds for the variable  $\xi^{(0)}$  subjected to the regression analysis (see formula (138))

$$(165) \quad \xi^{(0)} = \eta^{(n)} + a_1 \xi^{(1)} + a_2 \xi^{(2)} + \dots + a_n \xi^{(n)}.$$

Considering, on the other hand, the residuals  $\eta(t; k)$  defined by (153) and obtained from a stationary process  $\{\xi(t)\}$ , it may be (cf. (148)) that the minimizing coefficients  $a(i, k)$  are bounded in modulus by a constant not surpassing  $1/\alpha$ . In such a case, a diagonal selection procedure will show that there exist a real sequence  $a_1, a_2, \dots$  and a sequence of integers  $k_1, k_2, \dots$  such that for all  $i$

$$\lim_{s \rightarrow \infty} a(i, k_s) = a_i, \quad |a_i| \leq 1/\alpha.$$

Hence, we are led to ask if for these coefficients  $a_i$

$$\lim_{n \rightarrow \infty} [\eta(t) + a_1 \xi(t-1) + \dots + a_n \xi(t-n)]$$

exists and equals  $\xi(t)$ , a relation which would correspond to (165). If the answer is affirmative, a non-singular stationary process  $\{\xi(t)\}$  could always be written on the form

$$(166) \quad \{\eta(t)\} + a_1 \{\xi(t-1)\} + a_2 \{\xi(t-2)\} + \dots,$$

where  $\{\eta(t)\}$  is non-autocorrelated, and  $\eta(t)$  is uncorrelated with  $\xi(t-1)$ ,  $\xi(t-2)$ , etc.

However, as will be seen in section 26, certain conditions would have to be imposed upon  $\{\xi(t)\}$  in order to secure the representation (166). For the present we shall leave open all questions in this matter, and proceed to an aspect of the finite case suggesting another canonical form of the stationary process.

Writing

$$\vartheta(t; k) = \xi(t) - \eta(t; k),$$

it is evident that the sequence  $\{\vartheta(t; k)\}$  is convergent as  $k \rightarrow \infty$ . Denoting the limit process by  $\{\vartheta(t)\}$ , we obtain

$$\{\xi(t)\} = \{\eta(t)\} + \{\vartheta(t)\}.$$

Since  $\vartheta(t; k)$  is a linear expression in  $\xi(t-1), \dots, \xi(t-k)$ , and thus uncorrelated with  $\eta(t+n)$  for all  $n \geq 0$ , it follows that  $\eta(t)$  is uncorrelated with the variables  $\vartheta(t)$ ,  $\vartheta(t-1)$ , etc. So far there is a complete analogy with the finite relation (165). In further analogy,  $\vartheta(t; k)$  can be written as a sum involving the uncorrelated residuals  $\eta(t-1; k-1), \dots, \eta(t-k+1; 1)$ , and  $\eta(t-k; 0) = \xi(t-k)$ . This circumstance suggests the question of whether  $\{\vartheta(t)\}$  is a linear expression in  $\{\eta(t-1)\}, \{\eta(t-2)\}$ , etc. Were the answer in the affirmative, then  $\{\xi(t)\}$  could always be written on the form  $\{\eta(t)\} + b_1 \{\eta(t-1)\} + b_2 \{\eta(t-2)\} + \dots$ , where  $\{\eta(t)\}$  is non-autocorrelated. It will be found that such a sum will not be sufficient as a canonical form for the stationary process — in general, a singular process  $\{\psi(t)\}$ , which is uncorrelated with  $\{\eta(t)\}$ , has to be added in order to obtain  $\{\xi(t)\}$ .

After these introductory remarks, let  $\{\xi(t)\}$  represent an arbitrary non-singular stationary process with finite dispersion  $\sigma$ , and with zero for mean value. In the first place, let an approximation procedure be performed on  $\xi(t)$  by means of the residuals  $\eta(t; k), \dots, \eta(t-n; k)$  given by (158). Let the new residuals be denoted  $\psi(t; n; k)$  and written

$$(167) \quad \psi(t; n; k) = \xi(t) - b(0; n; k) \cdot \eta(t; k) - \cdots - b(n; n; k) \cdot \eta(t - n; k).$$

Since  $\xi(t)$  is non-singular, the coefficients  $b$  minimizing  $D(\psi(t; n; k))$  will be uniquely determined.

The processes  $\{\psi(t; n; k)\}$  defined by (167) are, like  $\{\eta(t; k)\}$ , of type (61). The processes  $\{\psi(t; n; k)\}$  will next be subjected to a repeated passage to the limit, first in respect of  $k$ , and then in respect of  $n$ .

Letting  $k$  tend to infinity, and paying regard to the relation

$$\lim_{k \rightarrow \infty} r(\eta(t-p; k), \eta(t-q; k)) = 0,$$

which according to theorem 6 is valid for  $p \neq q$ , we get (cf. (163))

$$(168) \quad \lim_{k \rightarrow \infty} b(p; n; k) = r(\xi(t), \eta(t-p)) / \alpha = b_p,$$

independently of  $n$ . Thus, keeping in mind that the variable  $\eta(t, t-1, \dots, t-n; k)$  tends to  $\eta(t, t-1, \dots, t-n)$  as  $k \rightarrow \infty$ , it follows that for all  $n$  and  $t$

$$(169) \quad \lim_{k \rightarrow \infty} \psi(t; n; k) = \xi(t) - \eta(t) - b_1 \eta(t-1) - \cdots - b_n \eta(t-n).$$

Let the limit variables thus obtained be denoted by  $\psi(t; n)$ . Now, holding  $n$  fixed, the variables  $\psi(t; n)$  will obviously constitute a stationary process  $\{\psi(t; n)\}$  which is the limit of the sequence  $\{\psi(t; n; 1)\}, \{\psi(t; n; 2)\}, \dots$

Keeping in mind that the variables  $\eta(t)$  are mutually uncorrelated, a short calculation shows that

$$(170) \quad D^2(\psi(t; n)) = [1 - (1 + b_1^2 + \cdots + b_n^2) \cdot \alpha^2] \sigma^2.$$

Concluding that the series  $\sum b_i^2$  is convergent, let us write

$$(171) \quad K^2 = 1 + b_1^2 + b_2^2 + \dots,$$

and further

$$(172) \quad \zeta(t; n) = \eta(t) + b_1 \cdot \eta(t-1) + \cdots + b_n \cdot \eta(t-n).$$

Thus prepared, let  $n$  tend to infinity in (169). Since  $\sum b_i^2$  converges, we have

$$D^2(\zeta(t; n+p) - \zeta(t; n)) = (b_{n+1}^2 + \cdots + b_{n+p}^2) \cdot D^2(\eta) \rightarrow 0$$

uniformly in  $p$  as  $n \rightarrow \infty$ . It follows that the sum  $\eta(t) + b_1 \eta(t-1) + \dots + b_p \eta(t-p) + \dots$  is convergent. Denoting the sum variable by  $\zeta(t)$ , we have

$$(173) \quad \zeta(t) = \lim_{n \rightarrow \infty} \zeta(t; n) = \eta(t) + b_1 \eta(t-1) + b_2 \eta(t-2) + \dots$$

Further, the variables  $\zeta(t)$  and  $\psi(t) = \lim_{n \rightarrow \infty} \psi(t; n)$  constitute two stationary processes  $\{\zeta(t)\} = \lim_{n \rightarrow \infty} \{\zeta(t; n)\}$ , and  $\{\psi(t)\} = \lim_{n \rightarrow \infty} \{\psi(t; n)\}$  respectively.

Observing that (173) yields

$$(174) \quad D^2(\zeta(t)) = (1 + b_1^2 + b_2^2 + \dots) \cdot \kappa^2 \sigma^2 = \kappa^2 \cdot K^2 \cdot \sigma^2,$$

two cases may be distinguished:

(A)  $\kappa \cdot K = 1$ . Then  $D(\psi(t)) = 0$ , and

$$(175) \quad \{\xi(t)\} = \{\zeta(t)\};$$

(B)  $\kappa \cdot K < 1$ . Then  $D(\psi(t)) > 0$ , and

$$(176) \quad \{\xi(t)\} = \{\zeta(t)\} + \{\psi(t)\}.$$

Advancing that  $\{\psi(t)\}$  is singular, it is seen that (176) covers both (175) and the cases II and III (see p. 81). Moreover, giving  $\psi(t)$  the same mean as  $\xi(t)$ , the representation (176) evidently holds also in case  $\{\xi(t)\}$  has a non-vanishing mean.

Formula (176) is the desired canonical form for a stationary process with finite dispersion. As already pointed out, the variable  $\zeta(t)$  corresponds directly with the case of a variable  $\xi$  in a finite number of dimensions. Further, according to (164) and (173), our  $\{\zeta(t)\}$  presents a certain similarity to the general process  $\{\gamma(t)\}$  of linear regression as introduced in section 15 γ. However,  $\{\zeta(t)\}$  is still more general, for the variables  $\eta(t)$  constituting  $\zeta(t)$  are non-correlated, while those forming  $\gamma(t)$  moreover are independent.

Some characteristic properties of the variables  $\zeta(t)$  and  $\psi(t)$  appearing in the canonical formula (176) will now be proved, and the main results then comprehended in a theorem.

Observing that

$$\left( \sum_{i=0}^{\infty} b_{p+i} \cdot b_i \right)^2 \leq \left( \sum_{i=0}^{\infty} b_{p+i}^2 \right) \cdot \left( \sum_{i=0}^{\infty} b_i^2 \right) = K^2 \cdot \sum_p b_i^2 \rightarrow 0 \text{ as } p \rightarrow \infty,$$

we conclude that  $\sum_{i=0}^{\infty} b_i \cdot b_{i+p}$  is convergent for all  $p$ , and that

$$(177) \quad r_p(\zeta) = E[\zeta(t) \cdot \zeta(t-p)] / \kappa^2 \cdot K^2 \cdot \sigma^2 = \\ = \lim_{n \rightarrow \infty} (b_p + b_{p+1} \cdot b_1 + \cdots + b_n \cdot b_{n-p}) / K^2 \rightarrow 0 \text{ as } p \rightarrow \infty.$$

Paying regard to (163), we get further

$$(178) \quad r_p(\zeta) = (b_p + b_{p+1} \cdot b_1 + b_{p+2} \cdot b_2 + \cdots) / K^2 = \\ = \frac{1}{\kappa K} \cdot \lim_{n \rightarrow \infty} r(\xi(t), \eta(t-p) + b_1 \cdot \eta(t-p-1) + \cdots + b_n \cdot \eta(t-p-n)) = \\ = r(\xi(t), \zeta(t-p)) / \kappa \cdot K.$$

Considering in the second place  $\{\psi(t)\}$ , we obtain from (170)

$$(179) \quad D^2(\psi(t)) = (1 - \kappa^2 \cdot K^2) \cdot \sigma^2,$$

Next, in case  $D(\psi) > 0$  we get from (169)

$$r(\psi(t; n), \eta(t+p)) = r(\xi(t) - \eta(t) - b_1 \cdot \eta(t-1) - \cdots - b_n \cdot \eta(t-n); \eta(t+p)).$$

Keeping in mind that  $\eta(t)$  is uncorrelated with any  $\xi(t-p)$  and with any  $\eta(t \pm p)$  for  $p \neq 0$ , and paying regard to (163), a short calculation shows that  $r(\psi(t; n), \eta(t+p)) = 0$  for  $p \geq -n$ . It follows that for any  $p \geq 0$

$$r(\psi(t), \eta(t+p)) = \lim_{n \rightarrow \infty} r(\psi(t; n), \eta(t+p)) = 0.$$

Hence the fundamental relation

$$(180) \quad r(\psi(t), \zeta(t+p)) = \lim r(\psi(t), \zeta(t+p; n)) = 0, \quad p \geq 0,$$

which shows that the processes  $\{\psi(t)\}$  and  $\{\zeta(t)\}$  are non-correlated. Thus we have (cf. (59) and (119))

$$(181) \quad r_k(\xi) = \frac{D^2(\psi)}{D^2(\xi)} \cdot r_k(\psi) + \frac{D^2(\zeta)}{D^2(\xi)} \cdot r_k(\zeta).$$

For the preparation of the remaining proof of the singularity of  $\{\psi(t)\}$ , let it be observed that the non-correlation between  $\eta(t)$  and  $\eta(t+p)$  for  $p \neq 0$  implies that for any real number  $a_1$

$$\begin{aligned} D^2(\zeta(t; n) + a_1 \cdot \zeta(t-1; n)) &= D^2(\eta(t) + (a_1 + b_1) \cdot \eta(t-1) + \\ &+ (a_1 \cdot b_1 + b_2) \cdot \eta(t-2) + (a_1 \cdot b_2 + b_3) \cdot \eta(t-3) + \cdots + a_1 \cdot b_n \cdot \eta(t-n-1)) = \\ &= D^2(\eta(t) + B_1 \cdot \eta(t-1) + \cdots + B_{n+1} \cdot \eta(t-n-1)) = \\ &= (1 + B_1^2 + B_2^2 + \cdots + B_{n+1}^2) \kappa^2 \cdot \sigma^2 \geq \kappa^2 \cdot \sigma^2, \end{aligned}$$

the auxiliary constants  $B_m$  introduced being real. By the same argument we conclude that for any real  $a_1, a_2, \dots, a_p$

$$(182) \quad D^2(\zeta(t) + a_1 \cdot \zeta(t-1) + \dots + a_p \cdot \zeta(t-p)) = \\ = \lim_{j \rightarrow \infty} D^2(\zeta(t; n_j) + a_1 \cdot \zeta(t-1; n_j) + \dots + a_p \cdot \zeta(t-p; n_j)) \geq x^2 \cdot \sigma^2.$$

Considering now the variable  $\xi(t) + a_1 \cdot \xi(t-1) + \dots + a_p \cdot \xi(t-p)$ , and paying regard to (176) and (180), a short calculation shows that

$$(183) \quad D^2(\xi(t) + a_1 \cdot \xi(t-1) + \dots + a_p \cdot \xi(t-p)) = \\ = D^2(\psi(t) + a_1 \cdot \psi(t-1) + \dots + a_p \cdot \psi(t-p)) + \\ + D^2(\zeta(t) + a_1 \cdot \zeta(t-1) + \dots + a_p \cdot \zeta(t-p)).$$

However, an  $\varepsilon > 0$  being arbitrarily given, we know from (156) that there exists a number  $p(\varepsilon)$  and a real sequence, say  $a_1^*, a_2^*, \dots, a_p^*$ , such that the left member of (183) is less than  $x^2 \cdot \sigma^2 + \varepsilon$ . On the other hand, (182) shows that the second variance in the right member of (183) is not below  $x^2 \cdot \sigma^2$ . Hence it follows that

$$D^2(\psi(t) + a_1^* \cdot \psi(t-1) + \dots + a_p^* \cdot \psi(t-p)) \leq \varepsilon.$$

Since  $\varepsilon$  is arbitrary, this relation implies that  $\{\psi(t)\}$  is singular of a finite or infinite rank.

Summing up, we have the following theorem in which one of the variables  $\{\psi(t)\}$  and  $\{\zeta(t)\}$  may be vanishing.

*Theorem 7.* Denoting by  $\{\xi(t)\}$  an arbitrary discrete stationary process with finite dispersion, there exists a three-dimensional stationary process  $\{\psi(t), \zeta(t), \eta(t)\}$  with the following properties:

- (A)  $\{\xi(t)\} = \{\psi(t)\} + \{\zeta(t)\}$ .
- (B)  $\{\psi(t)\}$  and  $\{\zeta(t)\}$  are non-correlated.
- (C)  $\{\psi(t)\}$  is singular.
- (D)  $\{\eta(t)\}$  is non-autocorrelated.
- (E)  $\{\zeta(t)\} = \{\eta(t)\} + b_1 \cdot \{\eta(t-1)\} + b_2 \cdot \{\eta(t-2)\} + \dots$   
where  $b_n$  represent real numbers such that  $\sum b_n^2$  is convergent.

*Illustrations.* In order to illustrate the autoregression analysis, let us consider a normal stationary process  $\{\xi(t)\}$  as defined by the characteristic function (111). Assuming for the sake of formal simplicity that  $m = 0$ , and that  $\sigma = 1$ , we shall first investigate a sum variable  $\zeta[t]$  of type (62).

Writing  $\zeta(t, t-1, \dots, t-n) = [\zeta(t), \zeta(t-1), \dots, \zeta(t-n)]$ , we have by definition

$$\left\{ \begin{array}{l} \zeta(t) = \xi(t) + a_1 \xi(t-1) + \cdots + a_h \xi(t-h) \\ \zeta(t-1) = \xi(t-1) + a_1 \xi(t-2) + \cdots + a_h \xi(t-h-1) \\ \vdots \quad \vdots \\ \zeta(t-n) = \xi(t-n) + a_1 \xi(t-n-1) + \cdots + a_h \xi(t-n-h) \end{array} \right.$$

According to the introductory remarks in section 14, the characteristic function of  $\zeta(t, t-1, \dots, t-n)$ , say  $f_n^*(Z_t, Z_{t-1}, \dots, Z_{t-n}) = e^{-\frac{1}{2}Q_n^*}$ , will for all  $n \leq h$  be obtained from the characteristic function  $f_{n+h}(X_t, X_{t-1}, \dots, X_{t-n-h})$  by the substitution

$$\begin{aligned}
 X_t &= Z_t \\
 X_{t-1} &= a_1 Z_t + Z_{t-1} \\
 \cdot &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
 X_{t-h} &= a_h Z_t + a_{h-1} Z_{t-1} + \cdots + Z_{t-h} \\
 X_{t-h-1} &= a_h Z_{t-1} + \cdots + a_1 Z_{t-h} + Z_{t-h-1} \\
 \cdot &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
 X_{t-n} &= a_h Z_{t-n+h} + \cdots + a_1 Z_{t-n+1} + Z_{t-n} \\
 X_{t-n-1} &= a_h Z_{t-n+h-1} + \cdots + a_2 Z_{t-n+1} + a_1 Z_{t-n} \\
 \cdot &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
 X_{t-n-h+1} &= a_h Z_{t-n+1} + a_{h-1} Z_{t-n} \\
 X_{t-n-h} &= a_h Z_{t-n}
 \end{aligned}
 \tag{184}$$

We conclude that the distribution of  $\zeta(t, t-1, \dots, t-n)$  is normal, and that we obtain  $Q_n^*$  from  $Q_{n+h}$  by the substitution (184).

According to the substitution theory of quadratic forms, the matrix defining  $Q_n^*$  is obtained as follows (see e.g. G. KOWALEWSKI (1909), § 94). Writing  $A_n(\xi)$  for the matrix of  $Q_n$ , and  $B_{h,n}$  for the matrix of the substitution (184), we have

$$A_{n+h}(\xi) = \begin{pmatrix} 1 & r_1 & r_2 & \dots & r_{n+h} \\ r_1 & 1 & r_1 & \dots & r_{n+h-1} \\ r_2 & r_1 & 1 & \dots & r_{n+h-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n+h} & r_{n+h-1} & r_{n+h-2} & \dots & 1 \end{pmatrix}$$

$$B_{h,n} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ a_1 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_h & a_{h-1} & a_{h-2} & \dots & 1 & 0 & \dots & 0 \\ 0 & a_h & a_{h-1} & \dots & a_1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_h & a_{h-1} & \dots & a_1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & 0 & a_h & a_{h-1} \\ 0 & 0 & \dots & \dots & 0 & 0 & a_h \end{pmatrix}$$

In the first place we must form a product matrix consisting of  $n + h + 1$  rows and  $n + 1$  columns, and take for the  $i^{th}$  element in the  $j^{th}$  row the inner product of the  $j^{th}$  row in  $A_{n+h}(\xi)$  by the  $i^{th}$  column in  $B_{n,h}$ . Multiplying then  $B_{h,n}$  and the product matrix by columns, we arrive at the matrix required. Denoting this by  $A_n(\zeta)$ , and by  $M'$  the transposed of a matrix  $M$ , we have

$$A_n(\zeta) = B'_{h,n} \cdot A_{n+h}(\xi) \cdot B_{h,n}.$$

Forming in the same way the infinite matrix

$$(185) \quad A(\zeta) = B' \cdot A(\xi) \cdot B,$$

where

$$(186) \quad A(\xi) = \begin{pmatrix} 1, r_1, r_2, \dots \\ r_1, 1, r_1, \dots \\ r_2, r_1, 1, \dots \\ \vdots \vdots \vdots \vdots \vdots \\ \vdots \vdots \vdots \vdots \end{pmatrix}, \quad B = \begin{pmatrix} 1, 0, 0, 0, \dots \\ a_1, 1, 0, 0, \dots \\ a_2, a_1, 1, 0, \dots \\ \vdots \vdots \vdots \vdots \vdots \\ a_h, a_{h-1}, a_{h-2}, \dots, a_1, 1, 0, 0, \dots \\ 0, a_h, a_{h-1}, a_{h-2}, \dots, a_1, 1, 0, \dots \\ 0, 0, a_h, a_{h-1}, \dots, a_1, 1, \dots \\ \vdots \vdots \vdots \vdots \vdots \end{pmatrix},$$

it is readily verified that  $A_n(\zeta)$  equals the principal minor of order  $n + 1$  in  $A(\zeta)$ . We conclude that the distribution of  $\zeta[t]$  is normal, and that  $A(\zeta)$  is the matrix of the infinite quadratic form  $Q^*(Z_t, Z_{t-1}, \dots)$  appearing in the exponent of the characteristic function of  $\zeta(t, t-1, \dots)$ .

Formally, the above procedure applies even in the case of an infinite sequence  $(a_1, a_2, \dots)$ . It is seen that if the double series appearing in the matrix  $Q^*$  are absolutely convergent, the variable  $\zeta[t] = \sum_{i=0}^{\infty} a_i \xi[t-i]$  will be well-defined, and constitute a normal process. The characteristic function of the variable  $\zeta(t, t-1, \dots)$  will be given by  $e^{-\frac{1}{2} Q^*}$ .

Next we shall consider a few particular instances.

Let  $(1, b_1, b_2, \dots)$  represent a real sequence such that  $K^2 = 1 + \sum b_i^2$  is finite, and let  $\{\eta(t)\}$  be a normal and purely random process with vanishing mean, and dispersion equaling unity. Considering the variable  $\zeta[t] = \eta[t] + b_1 \eta[t-1] + b_2 \eta[t-2] + \dots$ , it is readily verified that the above substitution procedure gives  $A(\zeta) = B' \cdot A(\eta) \cdot B$ , where

$$A(\eta) = \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \dots \\ \vdots & \vdots & \vdots & \dots \end{pmatrix}, \quad B = \begin{pmatrix} 1, 0, 0, 0, \dots \\ b_1, 1, 0, 0, \dots \\ b_2, b_1, 1, 0, 0, \dots \\ b_3, b_2, b_1, 1, 0, \dots \\ \vdots \vdots \vdots \vdots \vdots \vdots \end{pmatrix}, \quad A(\zeta) = \begin{pmatrix} K^2 & K^2 r_1 & K^2 r_2, \dots \\ K^2 r_1 & K^2 & K^2 r_1, \dots \\ K^2 r_2 & K^2 r_1 & K^2, \dots \\ \vdots \vdots \vdots \vdots \vdots \vdots \end{pmatrix},$$

having in the latter matrix written  $r_k$  for  $(b_k + b_1 b_{k+1} + b_2 b_{k+2} + \dots) / K^2$ . These results are seen to be in full agreement with formula (177), and with the correlation properties of a normal stationary process (see p. 62).

Considering in the second place a process  $\{\xi(t)\}$  which is singular of rank  $h$ , there exists, by definition, a sequence  $(a_1, \dots, a_h)$  such that the relation (77) is satisfied. Accordingly, the infinite matrix  $A(\zeta) = B' \cdot A(\xi) \cdot B$  formed by means of the matrices (186), will consist entirely of zeros. For instance, letting  $\xi(t) + \xi(t-1) = 0$  be the relation of singularity, we have  $h = a_1 = 1$  (cf. (107)). A short calculation will show that

$$A(\xi) = \begin{pmatrix} 1, -1, & 1, -1, \dots \\ -1, & 1, -1, & 1, \dots \\ 1, -1, & 1, -1, \dots \\ -1, & 1, -1, & 1, \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 1 & 1 & 0 & 0 & \dots \\ 0 & 1 & 1 & 0 & \dots \\ 0 & 0 & 1 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}, \quad A(\zeta) = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

If  $\{\xi(t)\}$  is singular of infinite rank, there exists, for every integer  $n$  and every  $\varepsilon > 0$ , a number  $h(\varepsilon, n)$  and a sequence  $a_i(\varepsilon, n)$  such that the variable  $\xi[t] + a_1 \xi[t-1] + \dots + a_h \xi[t-h]$  will give rise to a matrix  $Q_n^*$ , whose elements are all less than  $\varepsilon$  in modulus.

Proceeding to the operation of summing independent processes, let  $\{\zeta(t)\}$  and  $\{\psi(t)\}$  stand for two independent normal processes. Denoting the sum process by  $\{\xi(t)\}$ , and indicating all symbols referring to the three processes by  $\xi$ ,  $\zeta$  and  $\psi$  respectively, and paying regard to the evident fact that the characteristic function of  $\xi(t, t-1, \dots)$  is the mathematical product of the characteristic functions of the variables  $\zeta(t, t-1, \dots)$  and  $\psi(t, t-1, \dots)$ , we obtain  $D^2(\xi) \cdot Q(\xi) = D^2(\zeta) \cdot Q(\zeta) + D^2(\psi) \cdot Q(\psi)$ , i.e.

$$D^2(\xi) \cdot \sum_{0 \ 0}^{\infty \ \infty} r_{|p-q|}(\xi) \cdot X_{t-p} X_{t-q} = D^2(\zeta) \cdot \sum_{0 \ 0}^{\infty \ \infty} r_{|p-q|}(\zeta) \cdot X_{t-p} X_{t-q} + D^2(\psi) \cdot \sum_{0 \ 0}^{\infty \ \infty} r_{|p-q|}(\psi) \cdot X_{t-p} X_{t-q}.$$

In full agreement with (59) we obtain the relation (181), and observe that mutually uncorrelated normal processes are always independent.

The simple types of normal process mentioned above are sufficient to illustrate theorem 7. Starting from a purely random normal process  $\{\eta(t)\}$  with suitable dispersion, forming a sum process  $\{\zeta(t)\}$  of type  $\{\eta(t)\} + b_1 \{\eta(t-1)\} + b_2 \{\eta(t-2)\} + \dots$ , and adding an independent normal process  $\{\psi(t)\}$  ruled by an appropriate singularity, we shall arrive at an arbitrarily prescribed normal process.

## CHAPTER III.

### On the theory of some special stationary processes.

#### 21. On the concept of stochastical difference equation.

The relation

$$(187) \quad \{\xi(t)\} + a_1 \cdot \{\xi(t-1)\} + \cdots + a_h \cdot \{\xi(t-h)\} = \{\eta(t)\}$$

arrived at in section 15 δ presents a formal analogy with an ordinary, or functional, difference equation

$$(188) \quad x(t) + a_1 \cdot x(t-1) + \cdots + a_h \cdot x(t-h) = y(t).$$

We have found under certain conditions concerning the coefficients  $a_i$ , and in case the process  $\{\eta(t)\}$  is purely random, that there exists a stationary process  $\{\xi(t)\}$  which satisfies (187) and is of type

$$(189) \quad \{\xi(t)\} = \{\eta(t)\} + b_1 \cdot \{\eta(t-1)\} + b_2 \cdot \{\eta(t-2)\} + \dots$$

On the other hand, since under general conditions a solution of (188) will be of the form

$$(190) \quad x(t) = y(t) + b_1 \cdot y(t-1) + b_2 \cdot y(t-2) + \dots,$$

a clear formal analogy may be seen between (187) and (188) also in respect of the solutions.

Expressing the situation in words, a solution of type (190) of a functional difference equation is a moving average performed on the function  $y(t)$  in the right member. Correspondingly, any sample series  $(\xi_t, \xi_{t-1}, \dots)$  connected with a process  $\{\xi(t)\}$  of type (189) may be looked upon as a moving average of a purely random series  $(\eta_t, \eta_{t-1}, \dots)$ .

Because of this parallelism, I propose to call (187) a *stochastical difference relation* between the processes  $\{\xi(t)\}$  and  $\{\eta(t)\}$ . If  $\{\eta(t)\}$  is known, and  $\{\xi(t)\}$  unknown, (187) will be termed a *stochastical*

*difference equation.* An interpretation in the language of the theory of oscillatory mechanisms will reveal some interesting connexions between functional and stochastical difference equations, and exemplify the wide applicability of the new concept.

An oscillatory mechanism presents certain intrinsic features relevant to the structure of the movement considered. Studying the movement in integral time points, these features are summed up in the relation

$$(191) \quad x(t) + a_1 \cdot x(t-1) + \cdots + a_h \cdot x(t-h) = 0.$$

Interpreting (191) as an ordinary difference equation, the solutions (see section 6) describe how the phenomenon would develop out from any initial values, say  $x(t-1) = x_{t-1}, \dots, x(t-h) = x_{t-h}$ , if there were no external influence present.

In the ordinary difference equations, the external factors are taken into account by means of the function  $y(t)$ . Thus, instead of the value  $-a_1 \cdot x_{t-1} - a_2 \cdot x_{t-2} - \cdots - a_h \cdot x_{t-h}$  to be expected for  $x(t)$  when the earlier values are known, the variable in question takes on the value  $y(t) - a_1 \cdot x_{t-1} - \cdots - a_h \cdot x_{t-h}$ . In this approach the external influence is dealt with as functional, i. e. uniquely determined at any future time point.

The stochastical approach differs in the allowance for the external influence upon the mechanism. Here the external factors are not dealt with as functionally determined; they are only assumed to be ruled by certain probability laws. These laws constitute the stochastical process  $\{\eta(t)\}$ ; in general, the probability laws are subjected only to the conditions (53)—(54) which express that the laws must not contradict themselves (cf. p. 3). The simple case investigated in section 15 δ corresponds to a purely random effect of the external factors, but nothing prevents us from approximating the external influence by a non-purely-random, or even a non-stationary process  $\{\eta(t)\}$ .

Having fixed a  $\{\eta(t)\}$ , any sample series  $(\dots, \eta_{t-1}, \eta_t, \eta_{t+1}, \dots)$  will describe an actual realization of the external development. Since the movement of the mechanism is known when the external factors are determined, the sample series  $(\dots, \eta_{t-1}, \eta_t, \eta_{t+1}, \dots)$  considered will correspond to a certain sample series, say  $(\dots, \xi_{t-1}, \xi_t, \xi_{t+1}, \dots)$ , of the process  $\{\xi(t)\}$ . However, as we possess only probability knowledge of the actual path  $(\dots, \eta_{t-1}, \eta_t, \eta_{t+1}, \dots)$  of the external factors, we can reach only probability laws about the

behaviour of the oscillatory mechanism. These probability laws concerning  $(\dots, \xi_{t-1}, \xi_t, \xi_{t+1}, \dots)$  constitute the process  $\{\xi(t)\}$ , and form a solution of the stochastical difference equation.

By means of the probability laws found for the mechanism, it will be possible to give information as to the average behaviour of the phenomenon considered, i.e. as to the expectations referring to  $\{\xi(t)\}$ . It should be observed that we cannot say in advance that the conclusions as to the average behaviour will be identical to those drawn from the functional difference equation (191). Nevertheless, it has often been argued — more or less explicitly — that any intrinsic tendency of the mechanism to produce periodic oscillations will, on the average, give rise to a corresponding oscillation in the phenomenon when influenced by random shocks (see e.g. SIR G. WALKER (1931), p. 522, and R. FRISCH (1933), p. 202). However, the following analysis will show that there are important instances when such inference based on analogy is incorrect, qualitatively as well as quantitatively. For instance, let  $C \cdot q^t \cdot \cos(\lambda_1 \cdot t + \varphi)$  represent a solution of (191), i.e. a damped oscillation characteristic of the mechanism, and consider the simple case of purely random external shocks. Then, even if there is no other intrinsic tendency to oscillation present, a periodogram analysis for the search of the frequency  $\lambda_1$  will be more or less misleading. As will be shown in section 25, there is in general a systematic deviation between  $\lambda_1$  and the abscissa for which the expectation of the periodogram ordinate presents a maximum. It may even happen that there is no maximum at all in the neighbourhood of  $\lambda_1$ .

As soon as external influence cannot be considered free from random elements, the stochastical difference equation should be preferred to the functional equation (cf. the quotations from G. U. YULE (1927) in section 10). It is also obvious that the former embraces the latter as a special case, for any function  $y(t)$  may be interpreted as a singular random process. Thus, the stochastical difference equations seem to merit particular interest.

When omitting all dispensable conditions as to  $\{\eta(t)\}$  and the  $a_i$ 's, the solutions of the stochastical difference equations become of a very general type, and embrace fundamentally different classes of random process. Having already seen in section 15 δ that the solutions cover the stationary processes of linear autoregression, let us in the second place consider the special equation

$$(192) \quad \{\xi(t) - \{\xi(t-1)\}\} = \{\eta(t)\},$$

taking as before a purely random process for  $\{\eta(t)\}$ . The solutions of this equation are seen to form a type case of the *discrete homogeneous process* (see e. g. H. CRAMÉR (1937), Ch. VIII). In sharp contrast to the stationary processes, the oscillations here tend to increase in amplitude as time goes on. We may express this fact by saying that the homogeneous process is *evolutive* (cf. p. 1). In the particular case (192), the process  $\{\xi(t)\}$  cannot be assumed to have been in movement during an infinite past. Accordingly, and in contradistinction to the stationary case, the analysis of this equation generally has to be restricted to an interval of type  $(t_0 \leq t \leq \infty)$ .

As already pointed out, there are many problems calling for investigation in connexion with the general stochastical difference equation. The coming section is reserved for some groundwork concerning such equations. In accordance with the program of the present study, non-stationary solutions will be dealt with only very briefly.

## 22. Some fundamentals concerning stochastical difference equations.

According to the definition given in the previous section,

$$(193) \quad \{\xi(t)\} + a_1 \cdot \{\xi(t-1)\} + \cdots + a_h \cdot \{\xi(t-h)\} = \{\eta(t)\}$$

forms a stochastical difference equation in  $\{\xi(t)\}$  if the coefficients  $a_i$  are real, and if  $\{\eta(t)\}$  is a discrete random process. If  $a_h \neq 0$ , the equation will be termed of *order h*.

Let first an equation of order  $h$  with vanishing right member be considered,

$$(194) \quad \{\xi(t)\} + a_1 \cdot \{\xi(t-1)\} + \cdots + a_h \cdot \{\xi(t-h)\} = 0.$$

If there are any solutions to this equation, these will be singular in the sense indicated in section 14, and have (77) with  $m=0$  for relation of singularity. Being particularly interested in stationary solutions with finite dispersion, it follows from the analysis in section 14 that there exists a non-vanishing stationary process which satisfies (194) if, and only if, the characteristic equation (34) has at least one root on the circumference of the unit circle.

It is seen that if  $\{\xi(t)\}$  and  $\{\psi(t)\}$  are independent processes such that  $\{\xi(t)\}$  is a solution of (193), while  $\{\psi(t)\}$  satisfies (194),

then  $\{\xi(t)\} + \{\psi(t)\}$  will satisfy (194). This property of the stochastical difference equation forms another analogy to the functional case.

Secondly, we shall touch upon the case when the variable  $\{\eta(t)\}$  appearing in (193) is of the type  $\{\eta(t_0; t)\} = [\dots, 0, 0, \eta(t_0), \eta(t_0+1), \eta(t_0+2), \dots]$  considered in connexion with theorem 1. It is evident that this equation has one, and only one, solution of type  $\{\xi(t_0; t)\} = [\dots, 0, 0, \xi(t_0), \xi(t_0+1), \dots]$ , and that this solution is given by the following system,

$$\left\{ \begin{array}{l} \xi(t_0) = \eta(t_0), \\ \xi(t_0+1) = \eta(t_0+1) - a_1 \cdot \xi(t_0) = \eta(t_0+1) - a_1 \cdot \eta(t_0), \\ \xi(t_0+2) = \eta(t_0+2) - a_1 \cdot \xi(t_0+1) - a_2 \cdot \xi(t_0) = \eta(t_0+2) - a_1 \cdot \eta(t_0+1) + \\ \quad + (a_1^2 - a_2) \cdot \eta(t_0), \\ \dots \end{array} \right.$$

The general variable  $\xi(t_0 + t)$  is evidently of type

$$(195) \quad \xi(t_0 + t) = \eta(t_0 + t) + b_1 \cdot \eta(t_0 + t - 1) + b_2 \cdot \eta(t_0 + t - 2) + \dots + b_t \cdot \eta(t_0).$$

The coefficients  $b_i$  introduced are seen to be identical to those used in section 15, §. Thus,  $b_1, b_2, \dots, b_h$  will be obtained from the system (97), and the following ones from the difference equation (96). The following elementary theorem concerning the coefficients  $b_i$  will prove useful.

*Theorem 8.* *The series  $b_t$  defined by (96) and (97) does not satisfy any difference equation of type (32) of lower order than  $h$ .*

Writing  $b_t$  on the form (33), let us examine the values, say  $b(t)$ , of this analytical function taken on for  $t \leq 0$ . Since every linear difference relation which is satisfied by this function for  $t > 0$  must hold also for  $t \leq 0$  and *vice versa*, it is sufficient to verify theorem 8 for  $t \leq 0$ . According to the difference equation (96), we get  $m=0$  and

$$(196) \quad \left\{ \begin{array}{l} b_1 + a_1 \cdot b(0) + a_2 \cdot b(-1) + \dots + a_{h-1} \cdot b(-h+2) + a_h \cdot b(-h+1) = 0 \\ b_2 + a_1 \cdot b_1 + a_2 \cdot b(0) + \dots + a_{h-1} \cdot b(-h+3) + a_h \cdot b(-h+2) = 0 \\ \dots \\ b_{h-1} + a_1 \cdot b_{h-2} + a_2 \cdot b_{h-3} + \dots + a_{h-1} \cdot b(0) + a_h \cdot b(-1) = 0 \\ b_h + a_1 \cdot b_{h-1} + a_2 \cdot b_{h-2} + \dots + a_{h-1} \cdot b_1 + a_h \cdot b(0) = 0. \end{array} \right.$$

Identifying the left members of the  $h^{\text{th}}$  equations in the two systems (97) and (196), we obtain  $a_h \cdot b(0) = a_h$ . Since  $a_h \neq 0$ , this relation gives

$$(197) \quad b(0) = 1.$$

Inserting  $b(0) = 1$  in the system (196), the  $(h-1)^{\text{th}}$  equations of the two systems considered give  $a_h \cdot b(-1) = 0$ , or  $b(-1) = 0$ . In the same way we obtain successively

$$(198) \quad 0 = b(-1) = b(-2) = \cdots = b(-h+1).$$

Now, if  $b_t$  also satisfied a difference equation of lower order than  $h$ , it would follow from (198) that  $b(0) = 0$ , which contradicts (197).

We conclude from the theorem just proved that none of the individual components will be identically vanishing when  $b_t$  is written on the form (33). Thus, according to section 6, the two series  $\sum_{i=1}^{\infty} |b_i|$  and  $\sum_{i=1}^{\infty} b_i^2$  will be convergent if, and only if, all roots of the equation (34) are lying within the boundary of the unit circle. This corollary is important to the following.

Recurring to the solution  $\{\xi(t_0; t)\}$ , let the case be considered when the variables  $\eta(t)$  are independent, and have identical distribution functions. Then we obtain from (195)

$$D^2(\xi(t_0; t)) = D^2(\eta(t)) \cdot [1 + b_1^2 + b_2^2 + \cdots + b_{t-t_0}^2].$$

Since  $D(\eta_1) > 0$ , the process  $\{\xi(t_0; t)\}$  thus will be evolutive if the series  $\sum_{i=1}^{\infty} b_i^2$  is divergent. According to the above, divergence takes place if one or more of the roots of the characteristic equation are lying on the boundary of or outside the unit circle.

The process of linear autoregression is, by construction, a solution of a stochastical difference equation such that (A) the variable in the right member is purely random, and (B) all roots of the characteristic equation are of a modulus less than unity. Leaving aside the question of whether there are other equations with stationary solutions — it seems likely that certain equations involving a singular and stationary  $\{\eta(t)\}$  and where all the roots of the characteristic equation are of modulus unity are satisfied by suitable stationary and singular processes  $\{\xi(t)\}$  — this short introduction will be terminated by the proof of the following theorem, which shows that the condition (A) can be generalized in as much

as it is sufficient that the right member is stationary. In the language of the theory of oscillatory mechanisms, the theorem states that a mechanism whose intrinsic movements are damped will give rise to a stationary oscillation when influenced by external shocks of a stationary kind.

*Theorem 9.* Let (193) be a stochastical difference equation such that all roots of its characteristic equation are of a modulus less than unity, let  $\{\eta(t)\}$  be stationary and have a finite dispersion, and let the sequence  $b_1, b_2, \dots$  be given by (96) and (97). Then

$$\lim_{n \rightarrow \infty} [\{\eta(t)\} + b_1 \cdot \{\eta(t-1)\} + b_2 \cdot \{\eta(t-2)\} + \dots + b_n \cdot \{\eta(t-n)\}]$$

will exist, and form a stationary solution of the equation.

Observing that for any  $m > 0$

$$\begin{aligned} D^2(b_n \cdot \eta(t-n) + b_{n+1} \cdot \eta(t-n-1) + \dots + b_{n+m} \cdot \eta(t-n-m)) &\leq \\ &\leq (|b_n| + |b_{n+1}| + \dots + |b_{n+m}|)^2 \cdot D^2(\eta(t)) \rightarrow 0 \text{ as } n \rightarrow \infty, \end{aligned}$$

we conclude in the first place that

$$\lim_{n \rightarrow \infty} [\eta(t) + b_1 \cdot \eta(t-1) + \dots + b_n \cdot \eta(t-n)]$$

will exist. Denoting the limit variable by  $\xi(t)$ , an application of theorem 1 shows that the variables  $\xi(t)$  arrived at will constitute a stationary process  $\{\xi(t)\}$ . That this process satisfies (193), may be proved in the same way as the identity (100).

### 23. On the stationary processes with finite dispersion and with no singular component.

When writing a stationary process with finite dispersion on the canonical form (176) it may happen that the singular component is vanishing (cf. section 15 γ). In this section, we shall derive some general formulae concerning such processes. In the coming sections of the present chapter, we shall use these formulae, and the previous analysis of stochastical difference equations, for a detailed study of the processes of moving averages and of linear autoregression.

Let  $\{\zeta(t)\}$  and  $\{\eta(t)\}$  represent two stationary processes such that

$$(199) \quad \zeta(t) = \eta(t) + b_1 \cdot \eta(t-1) + b_2 \cdot \eta(t-2) + \dots$$

$$(200) \quad \eta(t) = \zeta(t) + a_1 \cdot \zeta(t-1) + a_2 \cdot \zeta(t-2) + \dots$$

where

(A)  $\{\eta(t)\}$  is non-autocorrelated,

(B)  $D(\eta(t)) > 0$  is finite,

(C)  $E[\eta(t)] = 0$ ,

(D) the sum  $\sum b_k^2$  is convergent.

Thanks to the convergence of  $\sum b_k^2$ , the formulae involving only the coefficients  $b_k$  will all have a real meaning. On the other hand, the coefficients  $a_k$  have been introduced in a purely formal way, and all questions concerning their existence will be left open for later treatment in connexion with the analysis of special cases. As a matter of fact, the expressions which involve the coefficients  $a_k$  will be used mainly as a formal comprehension of the special cases of moving averages and of linear autoregression.

Replacing  $t$  in (199) by  $t-1, t-2, \dots$ , and inserting in (200), we obtain the following relations between the coefficients  $a_i$  and  $b_i$ ,

$$(201) \quad a_k + b_1 \cdot a_{k-1} + \dots + b_{k-1} \cdot a_1 + b_k = 0, \quad k = 1, 2, \dots$$

If the set  $(a_i)$  is given, the set  $(b_i)$  thus will be uniquely determined, and vice versa. We obtain for the first few coefficients

$$(202) \quad \begin{cases} a_1 = -b_1, \\ a_2 = -b_2 + b_1^2, \\ a_3 = -b_3 + 2b_1b_2 - b_1^3, \\ a_4 = -b_4 + 2b_1b_3 - 3b_1^2b_2 + b_2^2 + b_1^4, \end{cases} \quad \begin{cases} b_1 = -a_1, \\ b_2 = -a_2 + a_1^2, \\ b_3 = -a_3 + 2a_1a_2 - a_1^3, \\ b_4 = -a_4 + 2a_1a_3 - 3a_1^2a_2 + a_2^2 + a_1^4. \end{cases}$$

Keeping the notations used in section 19, we write

$$(203) \quad K^2 = 1 + b_1^2 + b_2^2 + \dots,$$

and remark that the vanishing of the singular component implies that  $K = 1/\alpha$ . Thus we have

$$(204) \quad D(\zeta) = K \cdot D(\eta),$$

while the general formulae (163) and (177) reduce to

$$K \cdot r(\zeta(t+n); \eta(t)) = b_n,$$

$$(205) \quad r_k = r_k(\zeta) = (b_k + b_1 \cdot b_{k+1} + b_2 \cdot b_{k+2} + \dots) / K^2,$$

writing shortly  $r_k$  for  $r_k(\zeta)$ .

Next, we shall derive a fundamental set of relations between the sequences (a), (b) and (r). Multiplying the two identities

$$(206) \quad \zeta(t+s) = \eta(t+s) + b_1 \cdot \eta(t+s-1) + b_2 \cdot \eta(t+s-2) + \dots$$

$$(207) \quad \zeta(t) + a_1 \cdot \zeta(t-1) + a_2 \cdot \zeta(t-2) + \dots = \eta(t),$$

and forming the expectations of the resulting two members, we obtain in the case of a negative  $s$  in (206)

$$(208) \quad r_k + a_1 \cdot r_{k-1} + \dots + a_{k-1} \cdot r_1 + a_k + a_{k+1} \cdot r_1 + a_{k+2} \cdot r_2 + \dots = 0,$$

for all  $k > 0$ . In the same way, taking  $s \geq 0$ , we get

$$(209) \quad (1 + a_1 \cdot r_1 + a_2 \cdot r_2 + \dots) \cdot D^2(\zeta) = D^2(\eta),$$

and

$$(210) \quad r_k + a_1 \cdot r_{k+1} + a_2 \cdot r_{k+2} + \dots = b_k / K^2, \quad k \geq 0.$$

Using the relations (149)–(152), we shall next derive a set of forecast formulae. For this purpose, we shall consider the variable  $\zeta(t+k)$  as conditioned by the development of the process up to the time point  $t$  inclusive. Writing

$$(211) \quad Fc[\zeta(t+k)] = F_t[\zeta(t+k)],$$

where  $(C) = (\zeta(t-k) = \zeta_{t-k}, \eta(t-k) = \eta_{t-k}, k = 0, 1, 2, \dots)$ , we have first the formal relations

$$(212) \quad \begin{cases} \zeta_{t-k} = \eta_{t-k} + b_1 \cdot \eta_{t-k-1} + b_2 \cdot \eta_{t-k-2} + \dots; & k = 0, 1, 2, \dots \\ \eta_{t-k} = \zeta_{t-k} + a_1 \cdot \zeta_{t-k-1} + a_2 \cdot \zeta_{t-k-2} + \dots; & k = 0, 1, 2, \dots \end{cases}$$

Since the variables  $\eta(t)$  are uncorrelated, (150) gives the linear forecast

$$(213) \quad F_t[\zeta(t+k)] = b_k \cdot \eta_t + b_{k+1} \cdot \eta_{t-1} + b_{k+2} \cdot \eta_{t-2} + \dots; \quad k = 1, 2, \dots$$

As verified below, we have further

$$(214) \quad F_t[\zeta(t+k)] = -a_1 \cdot F_t[\zeta(t+k-1)] - a_2 \cdot F_t[\zeta(t+k-2)] - \cdots - a_{k-1} \cdot F_t[\zeta(t+1)] - a_k \cdot \zeta_t - a_{k+1} \cdot \zeta_{t-1} - a_{k+2} \cdot \zeta_{t-2} - \cdots; \quad k = 1, 2, \dots$$

This relation, which makes possible a successive calculation of the linear forecasts  $F_t$ , reduces to (213) if every  $F_t[\zeta(t+k-i)]$  is written on the form (214), and every  $\zeta_{t-i}$  is expressed in the values  $\eta_{t-i}$  by means of (212). In fact, for  $i \geq 0$  the coefficient of  $\eta_{t-i}$  then becomes

$$-a_1 \cdot b_{k+i-1} - a_2 \cdot b_{k+i-2} - \cdots - a_{k+i-1} \cdot b_1 - a_{k+i},$$

and according to (201) this expression equals  $b_{k+i}$ .

Alternatively, we may express the forecasts  $F_t$  in terms of the values  $\zeta_{t-i}$ . Writing

$$(215) \quad F_t[\zeta(t+k)] = f_{k,0} \cdot \zeta_t + f_{k,1} \cdot \zeta_{t-1} + f_{k,2} \cdot \zeta_{t-2} + \cdots,$$

we record in the first place

$$f_{1,i} = -a_i.$$

Further, inserting (215) in the left member of (214), and writing also the forecasts in the right member on the same form, we obtain

$$(216) \quad \begin{cases} f_{k,0} + a_1 \cdot f_{k-1,0} + a_2 \cdot f_{k-2,0} + \cdots + a_{k-1} \cdot f_{1,0} + a_k = 0, \\ f_{k,1} + a_1 \cdot f_{k-1,1} + a_2 \cdot f_{k-2,1} + \cdots + a_{k-1} \cdot f_{1,1} + a_{k+1} = 0, \\ \dots \dots \dots \dots \dots \dots \end{cases}$$

Thus, after having calculated the coefficients  $f_{k-i,j}$  appearing in the forecasts  $F_t[\zeta(t+k-i)]$ , the relations (216) yield the coefficients  $f_{k,j}$  necessary for computing  $F_t[\zeta(t+k)]$  in terms of the  $\zeta_{t-i}$ 's.

The relations (213)–(215) will be referred to as the *forecasting formulae*. The sample series sections  $(\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \dots)$  or (and)  $(\eta_t, \eta_{t-1}, \eta_{t-2}, \dots)$  being given, these formulae furnish the best linear forecast as to the future development of the series, viz. in the following sense.

A particular sample series section, say  $(C) = (\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \dots)$  being given, there exists a set of constants  $f_{k,0}(C), f_{k,1}(C), f_{k,2}(C), \dots$  minimizing the expectation

$$(217) \quad E[\zeta_C(t+k) - f_{k,0}(C) \cdot \zeta_t - f_{k,1}(C) \cdot \zeta_{t-1} - f_{k,2}(C) \cdot \zeta_{t-2} - \cdots]^2.$$

In general, the constants  $f_{k,i}(C)$  will depend on  $(C)$ , and differ from the  $f_{k,i}$ 's. Now, considering on the other hand the expectation

$$E[\zeta_C(t+k) - f_{k,0} \cdot \zeta_t - f_{k,1} \cdot \zeta_{t-1} - f_{k,2} \cdot \zeta_{t-2} - \dots]^2,$$

our  $f_{k,i}$ 's possess the property that the (weighted) average of this expression based on all sets  $(C)$  becomes a minimum. This minimum is

$$(218) \quad (1 + b_1^2 + b_2^2 + \dots + b_{k-1}^2) \cdot D^2(\eta),$$

a formula which shows clearly the scope of the forecast method under consideration. A forecast  $F_t[\zeta(t+k)]$  over  $k$  time units is decreasing in efficiency as  $k$  increases. As  $k \rightarrow \infty$ , the expression (218) tends to  $K^2 \cdot D^2(\eta) = D^2(\zeta(t))$ . In other words, for large  $k$ -values the forecast  $F_t[\zeta(t+k)]$  is approximately of the same efficiency as the trivial forecast  $E[\zeta(t+k)] = E[\zeta(t)] = 0$ .

If  $\{\eta(t)\}$  is purely random, it follows from (152) that we have

$$F_t[\zeta(t+k)] = E_C[\zeta(t+k)].$$

In this case, the coefficients  $f_{k,i}(C)$  appearing in (217) are independent of  $(C)$ ,

$$f_{k,i}(C) = f_{k,i}.$$

It is seen that (213)–(215) then give for every  $(C)$  a forecast which is the best one according to the principle of least squares.

#### 24. On the process of linear autoregression. General developments.

As already mentioned, we shall in the present section investigate in some detail the process of linear autoregression as dealt with in the sections 15 & 22. Denoting the process by  $\{\zeta(t)\}$ , the defining relation will be written

$$(219) \quad \{\zeta(t)\} + a_1 \{\zeta(t-1)\} + \dots + a_h \{\zeta(t-h)\} = \{\eta(t)\}.$$

It will be observed that the process  $\{\eta(t)\}$  need not be purely random; the following developments are valid under the broader assumption that the stationary process  $\{\eta(t)\}$  is non-autocorrelated. As before, we shall assume that  $E[\eta(t)] = E[\zeta(t)] = 0$ .

The formal developments given in the previous section are all valid in the present case. In fact, in the sequence  $a_i$  we have  $a_n = 0$  for  $n > h$ , so the serial developments are only apparently infinite. In order to arrive at more precise knowledge, we shall next consider these developments in some detail.

As to the  $b_t$ -series, we already know from the analysis in section 22 of the general stochastical difference equation that  $b_t$  in the present case is of type (33), and that the oscillations are damped, and that the series does not satisfy any linear difference equation of lower order than  $h$ .

Formula (209) gives

$$(220) \quad (1 + a_1 r_1 + \cdots + a_h r_h) \cdot D^2(\xi) = D^2(\eta).$$

As to the autocorrelation coefficients, we obtain from (208) and (210) the following three groups of relations.

$$(221) \quad \left\{ \begin{array}{l} r_k + a_1 \cdot r_{k-1} + a_2 \cdot r_{k-2} + \cdots + a_{h-1} \cdot r_{k-h+1} + a_h \cdot r_{k-h} = 0 \\ \vdots \\ r_{h+1} + a_1 \cdot r_h + a_2 \cdot r_{h-1} + \cdots + a_{h-2} \cdot r_3 + a_{h-1} \cdot r_2 + a_h \cdot r_1 = 0 \\ r_h + a_1 \cdot r_{h-1} + a_2 \cdot r_{h-2} + \cdots + a_{h-2} \cdot r_2 + a_{h-1} \cdot r_1 + a_h = 0 \end{array} \right.$$

$$(222) \quad \begin{cases} r_{h-1} + a_1 \cdot r_{h-2} + a_2 \cdot r_{h-3} + \cdots + a_{h-2} \cdot r_1 + a_{h-1} + a_h \cdot r_1 = 0 \\ \vdots \\ r_1 + a_1 + a_2 \cdot r_1 + \cdots + a_{h-2} \cdot r_{h-3} + a_{h-1} \cdot r_{h-2} + a_h \cdot r_{h-1} = 0 \end{cases}$$

$$(223) \quad \left\{ \begin{array}{l} 1 + a_1 \cdot r_1 + a_2 \cdot r_2 + \cdots + a_{h-2} \cdot r_{h-2} + a_{h-1} \cdot r_{h-1} + a_h \cdot r_h = 1/K^2 \\ r_1 + a_1 \cdot r_2 + a_2 \cdot r_3 + \cdots + a_{h-1} \cdot r_h + a_h \cdot r_{h+1} = b_1/K^2 \\ \vdots \\ r_k + a_1 \cdot r_{k+1} + a_2 \cdot r_{k+2} + \cdots + a_{h-1} \cdot r_{h+k-1} + a_h \cdot r_{h+k} = b_k/K^2 \end{array} \right.$$

The first group is given in the paper of SIR G. WALKER (1931) already referred to. We quote from the same paper the observations that (32) constitutes a difference equation satisfied by the  $r_k$ -series for  $k \geq h$ , that this relation is the same as that satisfied by the  $b_k$ -series, and that both series thus present damped oscillations of type (33).<sup>1</sup> SIR G. WALKER mentions further that the observa-

<sup>1</sup> K. STUMPFFF ((1936), p. 35) gives the relation  $b_k = r_k$ , which is correct only in case  $k = 1$ . STUMPFFF's deduction of the expectation relations (127) is based on this error.

tion that the  $b_k$ -series satisfies the difference equation (32) has already been given by G. U. YULE (1927) for the special case  $h=1$ .

The second group (222) contains  $h-1$  relations, and involves the autocorrelation coefficients  $r_1, r_2, \dots, r_{h-1}$ . Now, we obtain the coefficients  $r_1, \dots, r_{h-1}$  directly in terms of the coefficients  $a_i$  by solving the system (222). In fact, assuming that the equations are connected by a linear relation, say with coefficients  $A_1, \dots, A_{h-1}$ , we get  $0 = A_1 + a_h A_{h-1} = A_1 a_h + A_{h-1} = A_1(1 - a_h^2)$ ; since  $|a_h| < 1$  it follows  $A_1 = 0$ , and similarly  $A_2 = \dots = A_{h-1} = 0$ . The following  $r_k$ -values are obtained recurrently by means of the difference relations (221). In this connexion we give the following parallel to theorem 8.

*Theorem 10.* Let  $\{\zeta(t)\}$  be a process of linear autoregression of order  $h$ . Then the autocorrelation coefficients of  $\{\zeta(t)\}$  satisfy no difference equation of type (32) of lower order than  $h$ .

For the proof, let us write  $r_k$  on the form (33), and consider the values, say  $r(k)$ , taken on by this function for  $k = 0, -1, -2, \dots$ . According to (177), we have  $m = 0$ . Next, comparing the relation

$$r_h + a_1 \cdot r_{h-1} + \dots + a_{h-1} \cdot r_1 + a_h \cdot r(0) = 0$$

with the last equation in the system (221), we find  $a_h \cdot r(0) = a_h$ . Since  $a_h \neq 0$ , we conclude

$$(224) \quad r(0) = 1.$$

Comparing further the first equation in the system (222) with

$$r_{h-1} + a_1 \cdot r_{h-2} + \dots + a_{h-2} \cdot r_1 + a_{h-1} \cdot r(0) + a_h \cdot r(-1) = 0,$$

and paying regard to (224), we obtain  $r(-1) = r_1$ . This procedure may be continued  $h-1$  steps, which gives

$$(225) \quad r(-1) = r_1, \quad r(-2) = r_2, \quad \dots, \quad r(-h+1) = r_{h-1}.$$

Proceeding one step further, we arrive at the first equation in the system (223), and obtain  $a_h \cdot r(-h) + \frac{1}{K^2} = a_h \cdot r_h$ , from which we conclude

$$(226) \quad r(-h) \neq r_h.$$

Now, considering the function  $d_k$  defined for  $k = 0, \pm 1, \pm 2, \dots$  by

$$(227) \quad d_k = r_k - r(-k),$$

the relations (224)–(226) yield

$$(228) \quad d_{-h} \neq 0; \quad d_{-h+1} = d_{-h+2} = \dots = d_{-1} = d_0 = d_1 = \dots = d_{h-1} = 0; \\ d_h \neq 0.$$

On the other hand, it follows from (227) that  $d_k$  satisfies a difference equation of type (32) of an order twice that of the equation satisfied by  $r_k$ . However, by the argument used in theorem 8 it follows from (228) that  $d_k$  cannot possibly satisfy an equation (32) of lower order than  $2h$ . Thus  $r_k$  will satisfy no equation of type (32) of lower order than  $h$ , a reflection which completes the proof.

Among the general developments in section 23 there remains for consideration the forecasting formulae. Since  $a_k = 0$  for  $k > h$ , formula (214) shows that the forecasts  $F_t[\zeta(t+1)], F_t[\zeta(t+2)], \dots, F_t[\zeta(t+k)], \dots$ , concerning the time points  $t+1, t+2, \dots, t+k, \dots$ , forecasts based on the development up to the time point  $t$ , will satisfy the equation (32) in respect of  $k$ . Thus, the forecasts, too, will form a damped oscillation. Explaining in the terminology of the theory of oscillatory mechanisms, the forecast curve describes how the mechanism would develop out from the situation arrived at in the time point  $t$  if there were no external influence present in the future time points  $t+1, t+2, \dots$ . As by hypothesis the intrinsic oscillations of the mechanism are damped, the forecast curve will also be damped, in full agreement with the above. Thus,  $F_t[\zeta(t+k)] \rightarrow 0 = E[\zeta(t)]$  as  $k \rightarrow \infty$ , in agreement with the concluding remark of section 23.

Considering, finally, the relations (216), it will be observed that the coefficients  $f_{k,i}$  for all  $i$  satisfy the difference equation (32) in respect of  $k$ .

Next, we shall illustrate some points of the general analysis in Chapter II by means of the process of linear autoregression.

Keeping in mind that in the present case  $\sum_{k=0}^{\infty} |r_k|$  is convergent, we may apply the corollary to theorem 5 (see p. 69). We conclude that the generating function  $W(x)$  for all  $x$  in  $(0, \pi)$  has a bounded derivative given by (121). In order to transform this expression upon a finite form we write

$$G(x) = \sum_{k=0}^{\infty} r_k \cdot e^{ikx} = e^{isx} \cdot \sum_{t=-s}^{\infty} r_{t+s} \cdot e^{itx}.$$

Considering the identity

$$\begin{aligned} G(x) \cdot [e^{-ihx} + a_1 \cdot e^{-i(h-1)x} + \cdots + a_{h-1} \cdot e^{-ix} + a_h] &= \\ &= \sum_{t=-h}^{\infty} r_{t+h} \cdot e^{itx} + a_1 \cdot \sum_{t=-h+1}^{\infty} r_{t+h-1} \cdot e^{itx} + \cdots + \\ &\quad + a_{h-1} \cdot \sum_{t=-1}^{\infty} r_{t+1} \cdot e^{itx} + a_h \cdot \sum_{t=0}^{\infty} r_t \cdot e^{itx}, \end{aligned}$$

and paying regard to the relations (221), the right member reduces to

$$\begin{aligned} e^{-ihx} + (a_1 + r_1) \cdot e^{-i(h-1)x} + (a_2 + a_1 r_1 + r_2) \cdot e^{-i(h-2)x} + \cdots + \\ + (a_{h-1} + a_{h-2} r_1 + \cdots + r_{h-1}) \cdot e^{-ix}. \end{aligned}$$

Thus we obtain  $G(x)$  on the finite form

$$G(x) = \frac{1 + (a_1 + r_1) \cdot e^{ix} + \cdots + (a_{h-1} + a_{h-2} r_1 + \cdots + a_1 r_{h-2} + r_{h-1}) \cdot e^{i(h-1)x}}{1 + a_1 \cdot e^{ix} + a_2 \cdot e^{2ix} + \cdots + a_h \cdot e^{ihx}},$$

while  $W'(x)$  is given by

$$(229) \quad W'(x) = G(x) + G(-x) - 1 = 2 \cdot R[G(x)] - 1,$$

where  $R[G(x)]$  stands for the real part of  $G(x)$ .

Denoting by  $x_i$  the roots of (34), the roots of  $1 + a_1 x + \cdots + a_h x^h = 0$  are seen to equal  $1/x_i$ . Since these roots are lying outside the periphery of the unit circle, the denominator of  $G(x)$  is evidently non-vanishing, which is in full agreement with the earlier observation that  $W'(x)$  is bounded. Now, paying regard to the relations (127) and (129), and summing up the main results, we obtain the following theorem.

*Theorem 11. The generating function  $W(x)$  of the autocorrelation coefficients in a process  $\{\xi(t)\}$  of linear autoregression is absolutely continuous, and has a bounded derivative  $W'(x)$  given by (229). The expectation  $E[C^2(n; \lambda)]$  of an arbitrary ordinate in the SCHUSTER periodogram, as defined by (126), is given by*

$$E[C^2(n; \lambda)] = \frac{4 D^2(\xi)}{n} \cdot W'(\lambda) + O(1/n).$$

It should be noticed that the expectation is of the same order of magnitude as in the case of a purely random series (cf. (37)). When extending the analysed series, the periodogram ordinates thus tend to vanish.

In view of the applications of the theory of linear autoregression, it is of fundamental importance to investigate the possibilities of drawing conclusions from a sample series  $(\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \dots)$  upon the characteristic equation (191), or — interpreting in the language of the theory of oscillatory mechanisms — to investigate whether the past development of the mechanism as influenced by random external factors can give any information about its intrinsic oscillatory tendencies. In particular, is it possible to find out the periods of the intrinsic damped oscillations?

The previous analysis shows that the classical periodogram analysis is an inadequate method for the search of intrinsic oscillations — the longer the series analysed, the poorer the results. This conclusion holds both for the SCHUSTER and the WHITTAKER periodograms since they are of equal efficiency (see section 8). In the illustrations given in the next section, these periodogram questions are dealt with in more detail.

As has been emphasized by G. U. YULE (1927) and SIR G. WALKER (1931), an adequate tool for the search of the intrinsic properties of an oscillatory mechanism is yielded by the serial coefficients of the time series investigated. In fact, a serial coefficient  $\bar{r}_k$  approximates the corresponding autocorrelation coefficient  $r_k$ , and we know that the graph of  $r_k$  presents exactly those damped oscillations which are characteristic of the mechanism considered — there is conformity in respect to both the frequencies of the individual components of the oscillations and their damping exponential factors. Since a periodogram analysis is concerned with only the intrinsic frequencies, it is of particular importance that the serial coefficients can give information even about the damping factors.

Having above derived expressions for the autocorrelation coefficients and other characteristics connected with a process of linear autoregression, we are, in view of the applications, confronted with problems of an inverse type. In particular, it is seen that if the autocorrelation coefficients of a process of autoregression are known, the coefficients  $(a)$  will be obtained from the system (221—222). After having derived the coefficients  $(a)$  we obtain the primary process  $\{\eta(t)\}$  in terms of the process considered by means of form-

ula (200). Concluding that the inverse problem mentioned involves no difficulty in point of principle, reference is given to the next section and Chapter IV for illustrations.

A question now presenting itself concerns the reliability of the information yielded by the serial coefficients. Here we meet at once the same obstacles as in all significance problems concerning autocorrelated time series. In the first place, we notice that when forming the sampling variance of a serial coefficient, we arrive at a complicated expression involving i. a. an extensive sum of correlation coefficients between different serial coefficients. The difficulties of the problem having already been mentioned by G. U. YULE (1927), E. SLUTSKY (1934) has presented a large collection of formulae concerning the dispersion of various characteristics derived from sample series sections, formulae derived under the assumption that the variables considered are normally distributed.

However, it should be observed that the relevant problem does not consist merely in calculating the variance or the distribution of a single autocorrelation coefficient. We have also to face the much deeper question concerning the reliability of the periodicities which present themselves in the graph of the serial coefficients. In view of the complications already occurring in sampling problems involving merely one individual serial coefficient, the possibility of arriving at a practicable, quantitative measure of significance in this connexion seems, at least for the moment, hopeless.

There is also another essential difficulty. Correlation in time series, and correlation as considered in the classical applications, differ as to their quantitative significance (see H. WOLD (1936)). As a matter of fact, in the former case the correlation coefficients are, as a rule, quantitatively conditioned by the size of the statistical masses to which the coefficients refer. In order to give an example, we advance that certain business cycle data yield nice graphs of serial coefficients (see Chapter IV). For instance, the serial coefficients of the G. MYRDAL index 1830—1913 of the cost of living in Sweden show a clear damped oscillation (see e. g. fig. 14). Let it now be imagined that another index had been computed by the same method, covering the same period — 84 years —, but referring merely to a small part of Sweden. In the second case there would, of course, be a larger random element present in the index. A short reflection will show that the increase of the random element

is accompanied by a systematic tendency to a diminishing of the serial coefficients. — Consider, on the other hand, a classical application such as the correlation between cranial indices. Here the statistical units, the skulls, are uniquely determined, *unmodifiable* — in a material of say 84 skulls the correlation coefficient can be calculated in only one way. In other words, while we cannot possibly form a coefficient referring to a certain part of each of the 84 skulls, there is nothing illegitimate in the modifying of the 84 statistical units in the case of the time series correlation considered above. Advancing that the effect in question is treated in some detail in Appendix B, we point out for the present that in a theoretical autocorrelation model the size of the statistical mass must also be taken into consideration.

Summing up, the intricacy of the problems of significance, and the dependence of the correlation coefficients on the statistical mass, constitute two fundamental difficulties in quantitative autocorrelation analysis. The situation seems to justify the opinion that the utmost caution is necessary when drawing quantitative conclusions from observational time series — a hypothesis should not be considered safe unless corroborated by empirical series obtained from different and, if possible, independent statistical masses, and supported by aprioristic considerations independent of the statistical evidence.

The applications to observational data presented in Chapter IV are far from aiming at quantitative results, the purpose being more to exemplify the qualitative differences between the scheme of hidden periodicities and the schemes of linear regression. Since we attach no importance to the quantitative outcomes, the significance problems will not be entered upon in the present study.

## 25. On some special cases of linear autoregression.

Keeping the assumptions of the previous section, we shall in this section consider the special cases obtained when putting  $h = 1$  and  $h = 2$  in the general definition (219) of linear autoregression. The resulting formulae will be readily surveyable, and give rise to a few remarks of general scope. In a few instances the model series given in section 15 will be used for the illustrations.

Attaching the analysis to the formula (219), we take  $h = 2$ , and denote the roots of the characteristic equation (191) by  $p$  and  $q$ . Thus we have

$$(230) \quad a_1 = -(p + q), \quad a_2 = p \cdot q, \quad a_n = 0 \text{ for } n > 2; \quad |p| < 1, \quad |q| < 1;$$

$$\zeta[t] - (p + q) \cdot \zeta[t - 1] + p \cdot q \cdot \zeta[t - 2] = \eta[t].$$

As the coefficients  $a_i$  must be real, we notice that either I.  $p$  and  $q$  are real, or II.  $p = A + iB$ ,  $q = A - iB$ , where  $A$  and  $B$  represent real numbers such that

$$(231) \quad A^2 + B^2 = |p|^2 = |q|^2 < 1.$$

We shall assume that  $p \neq q$ , gaining thereby the general solution of the difference equation (32) to be (cf. also p. 146)

$$(232) \quad P_1 \cdot p^t + P_2 \cdot q^t,$$

where  $P_1$  and  $P_2$  are arbitrary. In case II, this expression may be written (cf. (33))

$$Q_1 \cdot C^t \cos \lambda_1 t + Q_2 \cdot C^t \sin \lambda_1 t.$$

where  $Q_1$  and  $Q_2$  are arbitrary, and

$$C = +\sqrt{A^2 + B^2}; \quad \cos \lambda_1 = A/C, \quad 0 < \lambda_1 < \pi.$$

Cases I and II will be dealt with separately.

### I. $p$ and $q$ are real.

Inserting the general expression (232) for  $b_1$  and  $b_2$  in the system (97), and solving for the constants  $P_1$  and  $P_2$ , we obtain readily

$$(233) \quad b_k = \frac{p}{p - q} \cdot p^k + \frac{q}{q - p} \cdot q^k = (p^{k+1} - q^{k+1})/(p - q), \quad k \geq 0.$$

Next, insertion of this expression in (203) yields

$$K^2 = \frac{D^2(\zeta)}{D^2(\eta)} = \frac{1 + pq}{(1 - p^2)(1 - q^2)(1 - pq)}.$$

The system (222) reduces to the single equation

$$(234) \quad r_1 + a_1 + a_2 \cdot r_1 = 0.$$

Solving for  $r_1$ , observing that  $r_0 = 1$  (cf. formula (224)), equalling these two coefficients to the general expression (232) for  $r_0$  and  $r_1$ , and solving the linear system thus obtained for the constants  $P_1$  and  $P_2$ , we get

$$(235) \quad r_k = \frac{p(1 - q^2)}{(p - q)(1 + pq)} \cdot p^k + \frac{q(1 - p^2)}{(q - p)(1 + pq)} \cdot q^k, \quad k \geq 0.$$

It is readily verified that (205) is satisfied. Using (230) and (235) in the formula (229), we find for the derivative of the generating function in the interval  $(0, \pi)$

$$(236) \quad W'(\lambda) = \frac{(1 - p^2)(1 - q^2)(1 - pq)}{(1 + pq)(1 + p^2 - 2p \cos \lambda)(1 + q^2 - 2q \cos \lambda)} = \\ = \frac{1}{K^2 \cdot (1 + p^2 - 2p \cos \lambda)(1 + q^2 - 2q \cos \lambda)}$$

and a short calculation will verify (121).

Theorem 11 gives for the expectation of an arbitrary ordinate in the SCHUSTER periodogram

$$E[C^2(n, \lambda)] = \frac{4D^2(\eta)}{n(1 + p^2 - 2p \cos \lambda)(1 + q^2 - 2q \cos \lambda)} + O(1/n).$$

The following special cases are instructive.

1)  $q = 0$ . In this case the relations reduce to

$$(237) \quad \zeta[t] - p\zeta[t-1] = \eta[t],$$

$$(238) \quad b_k = p^k, \quad D^2(\zeta) = D^2(\eta)/(1 - p^2), \quad r_k = p^k, \quad k \geq 0,$$

$$W'(\lambda) = \frac{1 - p^2}{1 + p^2 - 2p \cos \lambda} = \frac{D^2(\eta)}{D^2(\zeta) \cdot (1 + p^2 - 2p \cos \lambda)},$$

$$(239) \quad E[C^2(n, \lambda)] \cong \frac{4D^2(\eta)}{n \cdot (1 + p^2 - 2p \cos \lambda)}.$$

These formulae, which cover the case  $h = 1$ , have been given earlier by SIR G. WALKER ((1931), p. 521).<sup>1</sup>

<sup>1</sup> SIR G. WALKER gives also the variance formula (220) for an arbitrary  $h$  (formula K).

2)  $q = -p$ . We obtain without difficulty

$$\zeta[t] - p^2 \zeta[t-2] = \eta[t],$$

$$b_{2k} = r_{2k} = p^{2k}, \quad b_{2k+1} = r_{2k+1} = 0, \quad k \geq 0$$

$$D^2(\zeta) = D^2(\eta)/(1-p^4),$$

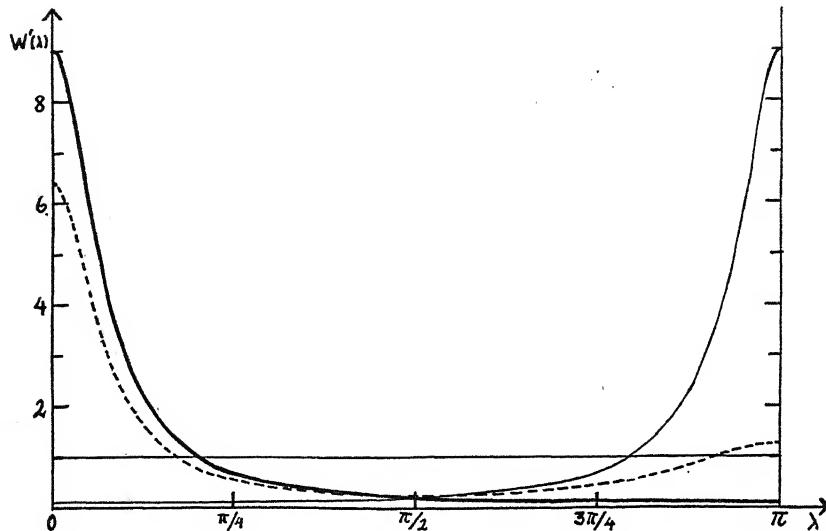


Fig. 2. Generating function derivatives obtained from formulae (236) and (239).

$$(240) \quad W'(\lambda) = \frac{1-p^4}{(1+p^2)^2 - 4p^2 \cdot \cos^2 \lambda},$$

$$E[C^2(n, \lambda)] = \frac{4 D^2(\eta)}{n \cdot [(1+p^2)^2 - 4p^2 \cdot \cos^2 \lambda]} + O(1/n).$$

The graph above shows the curves  $W'(\lambda)$  which belong to the processes defined by

- a) formula (103), i.e.  $p = -\cdot 8, q = 0$  (thin line),
- b) » (104), i.e.  $p = \cdot 8, q = 0$  (thick line),
- c)  $p = \cdot 8, q = -\cdot 6$  (broken line).

The graph contains, for comparison, the line  $W'(\lambda) = 1$  which represents the derivative of the generating function in the case of a purely random process (cf. the remark attached to formula (131)).

II.  $p$  and  $q$  are conjugate complex,

$$(241) \quad p = A + iB, \quad q = A - iB.$$

Since the above developments are perfectly general, we have now only to insert (241) in the formulae for  $b_k$ ,  $r_k$ , etc. given under I. After elementary transformations we get for  $k \geq 0$

$$\zeta[t] - 2A \cdot \zeta[t-1] + (A^2 + B^2) \cdot \zeta[t-2] = \eta[t],$$

$$(242) \quad b_k = C^k \cos k\lambda_1 + \frac{A}{B} \cdot C^k \sin k\lambda_1,$$

$$(243) \quad r_k = C^k \cos k\lambda_1 + \frac{A}{B} \cdot \frac{1 - C^2}{1 + C^2} C^k \sin k\lambda_1,$$

$$(244) \quad D^2(\zeta) = \frac{1 + C^2}{(1 - C^2)(1 + C^4 - 2A^2 + 2B^2)} \cdot D^2(\eta).$$

Writing

$$J(\lambda) = 4C^2 \cdot \left[ \frac{A(1 + C^2)}{2C^2} - \cos \lambda \right]^2 + \frac{B^2}{C^2}(1 - C^2)^2,$$

we get

$$(245) \quad W'(\lambda) = \frac{\lambda i^2(\eta)}{D^2(\zeta) \cdot J(\lambda)}$$

and

$$E[C^2(n, \lambda)] = \frac{4D^2(\eta)}{n \cdot J(\lambda)} + O(1/n).$$

Possessing now a collection of formulae sufficient for our purposes, a partial check is obtained by observing that when putting  $A = 0$  in the above expressions, we get the same result as when replacing  $p$  by  $i \cdot p$  in the formulae I, 2.

Proceeding to a first application of the above developments, we observe that the formulae under II cover the case of an oscillatory mechanism whose intrinsic oscillations consist of a single damped harmonic with frequency  $\lambda_1$  lying in the interval  $0 < \lambda_1 < \pi$ , and a damping factor  $C^t$ . Having earlier found that a periodogram analysis is ineffective in the case of linear autoregression, we are now in a position to prove the statement advanced in section 21 concerning the dangers of the periodogram method (see p. 95).

Speaking generally, the question is whether the expectation of the periodogram ordinate  $C^2(n, \lambda)$  presents a maximum when  $\lambda = \lambda_i$ , i. e. when the abscissa of the periodogram equals a frequency characteristic to the intrinsic oscillations of the mechanism. Were the answer in the affirmative, we should — at least in principle — be able to use a periodogram construction for the discovery of the intrinsic frequencies  $\lambda_i$ . In fact, considering a sample series  $(\zeta_{t-1}, \zeta_{t-2}, \dots)$ , making a sequence of periodograms on the basis of sections of type  $(\zeta_{t-1}, \dots, \zeta_{t-n}), (\zeta_{t-n-1}, \dots, \zeta_{t-2n}), \dots$ , and forming for every abscissa  $\lambda$  the average of the corresponding sequence of periodogram ordinates, the resulting curve would present maxima in the points  $\lambda = \lambda_i$  sought for. However, in order to prove that this way is barred — in point of principle, thus even if the statistical material were extensive enough for the construction of the required set of periodograms — we need only consider the case of a single intrinsic oscillation covered by the formulae (II) above. The expectation  $E[C^2(n, \lambda)]$  being asymptotically proportional to the derivative  $W'(\lambda)$ , it will be sufficient to investigate the extremes of the latter function.

Evidently,  $W'(\lambda)$  as given by (245) is maximized by those  $\lambda$ -values which minimize the auxiliary function  $J(\lambda)$ , and vice versa. The  $\lambda$ -values in question are seen to be

$$(246) \quad \lambda = \arccos \frac{A(1 + C^2)}{2C^2}, \quad \lambda = 0, \quad \lambda = \pi.$$

The behaviour of  $W'(\lambda)$  being different in the cases  $|A| \leq 2C^2/(1 + C^2)$ , the accompanying figure shows the part of the curve  $|A| = 2C^2/(1 + C^2)$  lying in the unit circle  $C = 1$ . An analysis of the second derivative of  $J(\lambda)$  gives the following results.

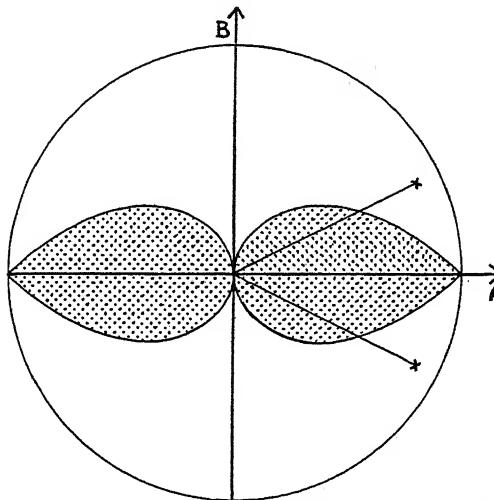


Fig. 3. Unit circle showing the domains where  $W'(\lambda)$  as given by (245) presents one or two maxima (non-dotted and dotted domains respectively)

a)  $|A| < 2C^2/(1+C^2)$ . Referring to the figure, the roots  $A \pm iB$  are in this case lying in the non-dotted part of the unit circle.

$W'(\lambda)$  presents one maximum in the interval  $(0, \pi)$ , viz. in the point  $\lambda$  defined by

$$(247) \quad \lambda = \arccos A(1 + C^2)/2C^2,$$

and two minima, viz. in the points  $\lambda = 0$ , and  $\lambda = \pi$ .

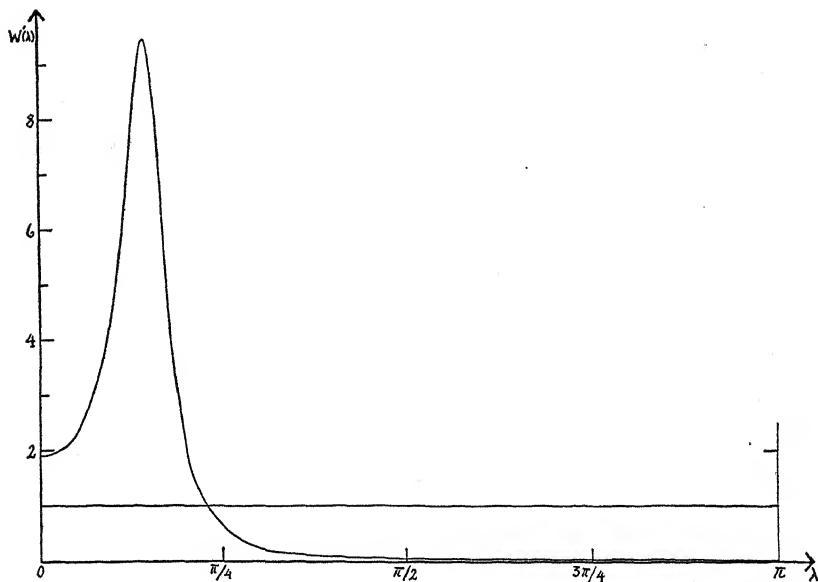


Fig. 4. Generating function derivative obtained from formula (245).

*Illustration.* Considering, for example, the process of type II obtained by taking  $A = .8$ ,  $B = .4$ , we get  $C^2 = .8$ ,  $A < 2C^2/(1+C^2)$ . In full agreement herewith, the point  $(A, B) = (.8, .4)$ , which is plotted in fig. 3, is lying in the non-dotted part of the unit circle. It follows from the above that the corresponding function  $W'(\lambda)$  presents but one maximum in the interval  $(0 \leq \lambda \leq \pi)$ . This function  $W'(\lambda)$  is shown in the figure above.

b)  $|A| \geq 2C^2/(1+C^2)$ . The roots are lying in the dotted part of the unity circle.

$W'(\lambda)$  presents one maximum and one minimum, the former being attained for  $\lambda = 0$  if  $A > 0$  and for  $\lambda = \pi$  if  $A < 0$ , the latter for  $\lambda = \pi$  if  $A > 0$  and for  $\lambda = 0$  if  $A < 0$ . The curves  $W'(\lambda)$  are here similar to those obtained in the case I, 1 (cf. fig. 2, unbroken lines).

Roughly speaking, if the roots  $A \pm iB$  of the characteristic

equation of the oscillatory mechanism are lying close to the periphery of the unit circle, the maximizing  $\lambda$ -value (247) is seen to approximate the intrinsic frequency  $\lambda_1 = \text{arc cos } A/C$ . In other words, if the intrinsic oscillations are only slightly damped, the periodogram analysis suggested above will be able to discover the frequency of the intrinsic oscillation.

On the other hand, holding  $A/B$  fixed, and letting  $C$  leave the immediate neighbourhood of the periphery of the unit circle, the  $\lambda$ -values resulting from (247) will deviate more and more from  $\lambda_1 = \text{arc cos } A/C$ . In case  $A > 0$ , the value given by (247) is seen to be less than  $\lambda_1$ , while the reverse holds true if  $A < 0$ . Thus, the periodogram will show a tendency to over-estimate the intrinsic period if this is above 4 time units, and to under-estimate it if the period is lying between 2 and 4 units. As to periods below 2 time units, these require a finer equidistance (cf. section 4). — We conclude next that the bias will be the larger, the more heavily the intrinsic oscillation is damped, i.e. the smaller the damping factor  $C$  is. Further, excepting the cases of an intrinsic period equalling exactly 2 or 4 time units, and letting the damping factor pass below a well-defined limit, the periodogram will altogether cease from giving information about the intrinsic period.

The disturbing effect pointed out may be looked upon as caused by the external factors influencing the oscillatory mechanism. As by hypothesis these factors contain a random element, I propose the term *chance effect* for the bias in question.

The situation may be described by saying that the inference drawn from the characteristic equation of the intrinsic oscillations does not apply directly to the oscillations of the mechanism when influenced by random external factors (cf. p. 95). While the chance effect is easily surveyed in the case of one intrinsic damped harmonic, the state of things is far more complicated when the mechanism presents a tendency to composite oscillations. Having stated this, our analysis of the chance effect will be brought to an end by a few explicit illustrations.

*Illustrations.* We shall first consider the case of one intrinsic oscillation, with damping factor  $C = \sqrt[4]{8} = .894$ , and frequency  $\lambda_1 = \text{arc tg } 1/2 = 26^\circ, 56$ . This case is elucidated by the diagrams 3 and 4. Inserting  $A = .8$  and  $C = \sqrt[4]{8}$  in (246), the  $\lambda$ -value maximizing  $W'(\lambda)$  is found to be  $\lambda = \text{arc cos } .9 = 25^\circ, 84$ . The periodogram thus tends to deliver a period equalling  $360/25.84 = 13.93$  time units, while the intrinsic period is  $13.55$  time units. In agreement with the general results concern-

ing the case of a non-composite intrinsic oscillation, the period is over-estimated. Although the damping factor is fairly large, just below unity, the bias is rather important.

Proceeding to some examples of periodogram ordinates derived from the model series given in section 15, let us first consider the process  $\{\delta^{(1)}(t)\}$  as defined by (103). Continuing in the notations of the present section, we obtain by short calculations  $h = 1$ ,  $p = .8$ ,  $D^2(\eta) = .6$ ,  $D^2(\zeta) = 1.667$ . Speaking in the language of oscillatory mechanisms, we are concerned with a single intrinsic harmonic, the frequency, period, and damping factor of which are  $\pi$ , 2, and .8 respectively. An ordinary periodogram analysis gives negative results, the expectance being inversely proportional to the length of the series analyzed. However, we know from the general theory that the expectation of the periodogram ordinate in  $\lambda = \pi$  is larger than the expectation in a purely random series. Taking  $n = 20$ , and keeping in mind that we are dealing with the exceptional case  $\lambda = \pi$ , the latter is found to equal  $1.667/n = .083$ . The former, as given by the exact formula (127) is  $E[C^2(20, \pi)] = .59$ . The proximate value obtained from (239) equals .75. Now, the 1000 elements in the model series  $(\delta_t^{(1)})$  have been arranged in 50 sections, each containing 20 consecutive elements, and the periodogram ordinate  $\bar{C}^2(20, \pi)$  has been computed for each section from formula (27). The average of the 50 ordinates thus obtained equals .56, a value not far from the the expectation .59.

Considering next the process  $\{\delta^{(2)}(t)\}$  given by (104) we have  $h = 1$ ,  $p = .8$ ,  $D^2(\eta) = .2$ ,  $D^2(\zeta) = .556$ . Referring again to Fig. 2, it is seen that for small frequencies the expectation is larger than in the case of a purely random series, but as before of the same order of magnitude in respect of  $n$ . Considering, e.g., the FOURIER coefficients (25) for  $k = 1$  in a sample series section of 20 elements, we have  $n = 20$ , and  $\lambda = 2\pi/n = 18^\circ$ . Formulae (127) and (239) give respectively  $E[C^2(20, 18^\circ)] = .36$  and  $E[C^2(20, 18^\circ)] \sim .34$ , while the corresponding value in the purely random case is .11. On the other hand, operating on the model series  $(\delta_t^{(2)})$  in the same way as before on  $(\delta_t^{(1)})$ , we have arrived at an average periodogram ordinate  $\bar{E}[\bar{C}^2(20, 18^\circ)]$  equalling .51. The rather substantial deviation from the expectation suggests that the periodogram ordinates are subject to a large dispersion. Actually, this suggestion seems to indicate how the matter stands. At any rate, the distributions of periodogram ordinates I have constructed on the basis of model time series have all presented a pronounced skewness, and a very large dispersion — often two or three of the largest sample values  $\bar{C}^2(n, \lambda)$  constitute alone as much as 20 to 30 per cent of the sum of the 50 sample values in the material. An instructive example of this is given below.

Recurring to the series  $(\delta_t^{(1)})$ , formula (127) yields  $E[C^2(20, 18^\circ)] = .046$ , while (239) gives  $E[C^2(20, 18^\circ)] \sim .038$ . A periodogram analysis as described in the previous illustration has given  $\bar{E}[\bar{C}^2(20, 18^\circ)] = .058$ . The figure below shows the distribution of the 50 averaged sample values of the periodogram ordinate dealt with. The figure contains the curve of summed relative frequencies ( $\bar{F}$ , thick lines), and a histogram showing the frequencies in the classes  $0 - .01$ ,  $.01 - .02$ , etc. ( $f$ , broken lines). The frequency curve graduating the histogram has been drawn by hand (broken curve). The distribution is seen to be very skew, and the histogram suggests no pronounced maximum for the frequency curve.

It is interesting to notice that the aforementioned investigation by E. SLUTSKY (1934) on sampling problems in the case of normally distributed time series data is

likewise discouraging with regard to the reliability of the periodogram. — As emphasized by SIR G. WALKER (see, e. g. D. BRUNT (1931) and R. A. FISHER (1929)), particular caution is in place when applying the periodogram method without the guidance of aprioristic knowledge of possible periods. In fact, when letting the largest ordinate indicate the frequency sought for, it should be taken into consideration that this ordinate has a larger expectance than a fixed ordinate, or a randomly chosen ordinate.

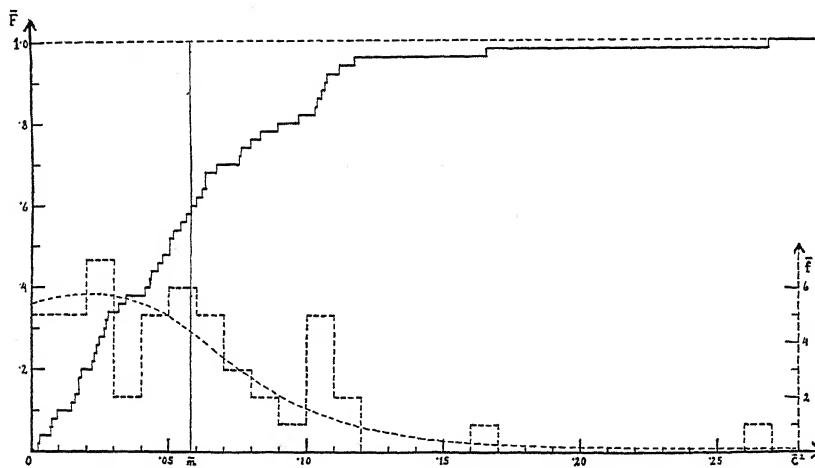


Fig. 5. Distribution of periodogram ordinates derived from the model series  $\bar{\delta}_t^{(1)}$  (see table 3).

Next, we shall give a few applications of the inversion formula (200). Working on the model series given in section 15 δ, we shall exemplify the construction of the primary series from a given series of autoregression.

*Illustrations.* Considering the model series  $(\bar{\delta}_t^{(1)})$  as defined by (103), we have  $\bar{\alpha}_t^{(2)} = \bar{\delta}_t^{(1)} + .8 \bar{\delta}_{t-1}^{(1)}$  for  $t = 1, 2, \dots$ . Inserting successively the  $\bar{\delta}$ -values given in table 3, we get  $\bar{\alpha}_1^{(2)} = 1.00$ ,  $\bar{\alpha}_2^{(2)} = .20 + .8 \times 1.00 = 1.00$ ,  $\bar{\alpha}_3^{(2)} = -.16 + .8 \times .20 = .00$ , etc., in full agreement with table 1. Thus, knowing the autoregression coefficient  $a_1 = .8$ , we can reconstruct the primary series of the autoregression series examined. In the same way, we can reconstruct without difficulty the primary series of the model series  $(\bar{\delta}_t^{(2)})$  and  $(\bar{\delta}_t^{(3)})$ .

In view of the applications, an important problem is to find the coefficients (a) belonging to a given series of autoregression. This problem will be discussed in detail in section 32. For the moment we shall only remark that the relations (222) together with the last relation in (221) form a system of linear equations which will permit us to derive the coefficients  $a_1, \dots, a_h$  in terms of the autocorrelation coefficients  $r_1, \dots, r_h$ . Thus, identifying the coefficients  $r_k$  with the serial coefficients  $\bar{r}_k$  of the series examined, the linear system will give us a set of proximate

coefficients  $a_1, \dots, a_h$ . We can then derive as before the primary series which corresponds to the coefficients  $a_1, \dots, a_h$  arrived at.

We give below the first five serial coefficients of the model series  $(\bar{\delta})$ , and the corresponding autocorrelation coefficients as derived from formulae (238) and (243).

Table 5. Serial and autocorrelation coefficients of the model series  $(\bar{\delta})$ .

	$k =$	1	2	3	4	5
$\bar{\delta}^{(1)}$	$\bar{r}_k =$	-·786	·647	-·498	·375	-·291
	$r_k =$	-·800	·640	-·512	·410	-·328
$\bar{\delta}^{(2)}$	$\bar{r}_k =$	·862	·700	·580	·482	·398
	$r_k =$	·800	·640	·512	·410	·328
$\bar{\delta}^{(3)}$	$\bar{r}_k =$	·127	-·628	-·194	·397	·186
	$r_k =$	·121	-·626	-·204	·366	·206

It is interesting to notice that although the model series consist of as many as 1000 elements each, the deviations between empirical and hypothetical correlation coefficients are rather substantial (cf. also p. 50 and p. 109).

Our analysis of the process of linear autoregression will be concluded by revealing a connexion with the »sinusoidal limit theorems» of E. SLUTSKY and V. ROMANOVSKY touched upon in section 16.

Let

$$L(x) = x^2 - 2Ax + 1 = 0, \quad -1 < A < 1,$$

stand for the characteristic equation of a simple harmonic

$$P_1 \cos \lambda_1 t + P_2 \sin \lambda_1 t,$$

and let  $\{\zeta^{(1)}(t)\}, \{\zeta^{(2)}(t)\}, \dots$  represent a sequence of processes of linear autoregression defined by

$$\zeta^{(p)}[t] - 2A_p \zeta^{(p)}[t-1] + C_p^2 \zeta^{(p)}[t-2] = \eta^{(p)}[t].$$

Let it further be assumed that the processes have equal dispersion,

$$D(\zeta^{(p)}(t)) = \sigma,$$

and that

$$\lim_{p \rightarrow \infty} A_p = A, \quad \lim_{p \rightarrow \infty} C_p = 1.$$

Attaching  $(p)$  to the symbols referring to the processes  $\{\zeta^{(p)}\}$ , we obtain from (204) and (244)

$$\sigma_p^2 = K_p^2 \cdot D^2(\eta^{(p)}) = \frac{1 + C_p^2}{(1 - C_p^2)(1 + C_p^2 - 2A_p^2 + 2B_p^2)} \cdot D^2(\eta^{(p)}).$$

Since by hypothesis  $1 - C_p^2$  tends to zero as  $p \rightarrow \infty$ , it follows that  $K_p \rightarrow \infty$  as  $p \rightarrow \infty$ . Concluding next from (242) that  $b_k^{(p)}$  is bounded, it follows that, for all  $k$ ,  $\lim_{p \rightarrow \infty} b_k^{(p)}/K_p^2 = 0$ . Considering now the systems (221)–(223), it is seen that

$$(248) \quad \lim_{p \rightarrow \infty} L(r_k^{(p)}) = 0, \quad -\infty < k < \infty.$$

The relation (248) thus arrived at implies that the sequence  $\{\zeta^{(p)}(t)\}$  is ruled by the sinusoidal limit theorem (cf. p. 64). Thus, approximating an arbitrary sample series  $(\zeta) = (\zeta_{t_0+1}^{(p)}, \dots, \zeta_{t_0+n}^{(p)})$  by a suitable simple harmonic with frequency  $\lambda_1$ , say  $x_p(t, \zeta)$ , and holding  $n$  fixed, it follows that for every  $\varepsilon > 0$

$$\lim_{p \rightarrow \infty} P[|x_p(t_0 + 1, \zeta) - \zeta_{t_0+1}^{(p)}| < \varepsilon, \dots, |x_p(t_0 + n, \zeta) - \zeta_{t_0+n}^{(p)}| < \varepsilon] = 1.$$

Since the proving arguments are of a general nature, they will be found to apply also in the case of a relation  $L(x) = 0$  of arbitrary order. The process of linear autoregression thus forms a convenient starting point for the construction of sequences covered by the sinusoidal limit theorems.<sup>1</sup> Obtaining in this way a continuous passage to a singular process, it is of interest to notice that the formula (199), together with the observation that  $\sum_{k=1}^{\infty} (b_k^{(p)})^2$  tends to infinity with  $p$ , forms a sufficient basis for proving that the singular limit process is normal.

## 26. On the process of moving averages.

Proceeding in the present section to an analysis of the process of moving averages as defined in section 15 β, we shall first apply the general formulae in section 23. The questions of when a representation of type (200) holds will give rise to a discussion elu-

<sup>1</sup> A sequence of moving averages cannot be ruled by the sinusoidal limit law unless the number of weights tends to infinity (cf. p. 65).

cidating certain points of the autoregression analysis as presented in sections 19 and 20.

By definition, the general formula for a process  $\{\zeta(t)\}$  of moving averages reads

$$(249) \quad \{\zeta(t)\} = \{\eta(t)\} + b_1\{\eta(t-1)\} + \cdots + b_h\{\eta(t-h)\},$$

where  $\{\eta(t)\}$  is purely random, and the sequence  $(b) = (b_1, \dots, b_h)$  is real. As before, we shall assume that  $D(\eta)$  is finite, and that  $E[\eta] = 0$ . Thus (249) forms a special case of the variable defined by (199), and it follows further that the formal developments remain valid if  $\{\eta(t)\}$  is non-autocorrelated.

In the process of linear autoregression, the autocorrelation coefficients and forecasting values were found to follow certain damped harmonics. In the present case only the first  $h$  elements in the sequences mentioned are different from zero. In fact, (204) and (205) yield for a process  $\{\zeta(t)\}$  given by (249)

$$D^2(\zeta) = (1 + b_1^2 + b_2^2 + \cdots + b_h^2) \cdot D^2(\eta),$$

$$(250) \quad r_k(\zeta) = \begin{cases} (b_k + b_1 \cdot b_{k+1} + \cdots + b_h \cdot b_{h-k}) / (1 + b_1^2 + \cdots + b_h^2) & \text{for } k \leq h, \\ 0 & \text{for } k > h, \end{cases}$$

where  $k \geq 0$  and  $b_0 = 1$ , formulae given by Professor H. CRAMÉR in his 1933 Course. Next, (213) gives

$$F_t[\zeta(t+k)] = \begin{cases} b_k \cdot \eta_t + b_{k+1} \cdot \eta_{t-1} + \cdots + b_h \cdot \eta_{t-h+k} & \text{for } 0 \leq k \leq h, \\ 0 & \text{for } k > h, \end{cases}$$

where the forecast is based upon the condition

$$(C) = (\eta(t) = \eta_t, \eta(t-1) = \eta_{t-1}, \dots, \eta(t-h+1) = \eta_{t-h+1}).$$

Inserting (250) in the general formula (116) for the generating function  $W(x)$  of the autocorrelation coefficients in a stationary process, we conclude that  $W(x) - x$  reduces to a finite trigonometrical sum. Thus, in the present case  $W'(x)$  exists, and is like  $W(x)$  uniformly bounded. Consequently, replacing »linear autoregression» by »moving averages» in theorem 11, we get a corresponding theorem covering the present case. We conclude, i. a., that periodogram analysis is an inadequate method of research in the case of moving averages also — the expectance of an arbitrary periodogram ordinate is inversely proportional to the length of the series under analysis. — Formula (256) gives  $W'(x)$  explicitly.

In view of the applications, the following questions call for investigation (cf. the illustrations on p. 119). The autocorrelation coefficients  $r_k$  of a process of moving averages being given, is it possible to derive the primary process and the coefficients ( $b$ ) of the moving average? In particular, does there exist a relation of type (200) yielding the primary process? If the answer to the second question is in the affirmative, how can the coefficients ( $a$ ) be obtained? The first of these problems was formulated by Professor H. CRAMÉR in his 1933 Course.

It is suitable to start the analysis with the second question. Denoting the process considered by  $\{\zeta(t)\}$ , we shall, until further notice, assume that  $\{\zeta(t)\}$  is given by (249).

In the first place we observe that theorem 9 covers the case when all the roots of the equation

$$(251) \quad x^h + b_1 x^{h-1} + \cdots + b_{h-1} x + b_h = 0$$

are of modulus less than unity. The theorem states that if this condition is satisfied, an infinite sequence  $(a) = (a_1, a_2, \dots)$  such that (200) holds is given by the system (97) and the difference relations obtained for  $a_i$  when replacing the  $a_i$ 's by the  $b_i$ 's in (96). These relations constitute a difference equation of order  $h$  satisfied by the sequence  $(a)$ , and it should further be observed that the sequence  $(a)$  arrived at forms a composed damped harmonic.

We have found that the condition attached to the roots of (251) will secure the representation (200). Under these circumstances the relations (208) and (210) derived from (206) and (207) must hold true. Explicitly, these read in the present case as follows:

$$(252) \quad a_{2h+k}r_h + a_{2h+k-1}r_{h-1} + \cdots + a_{h+k+1}r_1 + a_{h+k} + \\ + a_{h+k-1}r_1 + \cdots + a_{k+1}r_{h-1} + a_kr_h = 0,$$

$$(254) \quad \left\{ \begin{array}{l} a_h r_h + a_{h-1} r_{h-1} + \cdots + a_1 r_1 + 1 = 1/K^2 \\ a_{h-1} r_h + \cdots + a_1 r_2 + r_1 = b_1/K^2 \\ \vdots \\ a_2 r_h + a_1 r_{h-1} + r_{h-2} = b_{h-2}/K^2 \\ a_1 r_h + r_{h-1} = b_{h-1}/K^2 \\ r_h = b_h/K^2. \end{array} \right.$$

Comparing the above coefficients  $a_i$  and the coefficients  $a_{i,n} = -a(i,n)$  yielded by an autoregression analysis of  $\{\zeta(t)\}$  as described in section 19, and letting  $\varepsilon > 0$  be arbitrarily given, it follows that there exists an  $n$  such that

$$\begin{aligned} (\chi^2 + \varepsilon) D^2(\zeta) &\geq D^2(\zeta(t) + a_1 \zeta(t-1) + \cdots + a_n \zeta(t-n)) \geq \\ &\geq D^2(\zeta(t) + a_{1,n} \zeta(t-1) + \cdots + a_{n,n} \zeta(t-n)) \geq \chi^2 \cdot D^2(\zeta), \end{aligned}$$

the second inequality resulting from the definition of the coefficients  $a_{i,n}$ , and the third being similar to (182). Inserting (249), an elementary transformation yields

$$\begin{aligned} \varepsilon &\geq (a_1 + b_1)^2 + (a_2 + a_1 b_1 + b_2)^2 + \cdots + \\ &\quad + (a_n + a_{n-1} b_1 + \cdots + a_{n-h} b_h)^2 + \cdots + (a_n b_h)^2 \geq \\ &\geq (a_{1,n} + b_1)^2 + (a_{2,n} + a_{1,n} b_1 + b_2)^2 + \cdots + \\ &\quad + (a_{n,n} + a_{n-1,n} b_1 + \cdots + a_{n-h,n} b_h)^2 + \cdots + (a_{n,n} b_h)^2 \geq 0. \end{aligned}$$

Concluding that  $|a_{i,n} - a_i| < c_i \sqrt{\varepsilon}$ , where  $c_i$  is independent of  $n$ , it follows that  $a_{i,n} \rightarrow a_i$  as  $n \rightarrow \infty$ . Moreover, forming the difference of the residuals based on  $(a_1, \dots, a_n)$  and  $(a_{1,n}, \dots, a_{n,n})$ , a simple application of SCHWARZ' inequality shows that the non-autocorrelated process yielded by the autoregression analysis is identical with the primary process.

Considering the singular case when (251) presents at least one root on the boundary of the unit circle, we conclude from theorem 8 that the series  $\sum a_i^2$  will be divergent, and that no representation of type (200) will exist. By elementary transformations we shall show that this obstacle to calculating the primary process may be removed by means of a limit passage.

Let (249) represent a moving average,  $x_n$  the roots of the characteristic equation (251), and let  $|x_n| \leq 1$ . Introducing auxiliary averages  $\{\xi^{(i)}(t)\}$  defined by

$$\xi^{(i)}(t) = \eta(t) + B_1^{(i)} \cdot \eta(t-1) + \cdots + B_h^{(i)} \cdot \eta(t-h), \quad i = 1, 2, \dots,$$

let the symbols referring to  $\{\xi^{(i)}(t)\}$  be marked by  $(i)$ . Let  $x_n^{(i)} = x_n$  when  $|x_n| < 1$ , and let  $x_n - x_n^{(i)} = \Delta_i$ ,  $|x_n^{(i)}| = 1 - \varepsilon_i$  when  $|x_n| = 1$ , where  $0 < \varepsilon_i = |\Delta_i| \rightarrow 0$  as  $i \rightarrow \infty$ .

By construction, the inversion formula (200) applies to  $\{\xi^{(i)}(t)\}$ ,

$$\eta(t) = \xi^{(i)}(t) + a_1^{(i)} \cdot \xi^{(i)}(t-1) + a_2^{(i)} \cdot \xi^{(i)}(t-2) + \dots$$

The limit relation desired is based on the coefficients  $a_t^{(i)}$  and reads

$$(255) \quad \{\eta(t)\} = \lim_{i \rightarrow \infty} [\{\zeta(t)\} + a_1^{(i)} \cdot \{\zeta(t-1)\} + a_2^{(i)} \cdot \{\zeta(t-2)\} + \dots].$$

Writing  $\{\eta(t, i)\}$  for the process under the limes sign, it is sufficient to prove that  $D_i^2 = D^2(\eta(t, i) - \eta(t))$  is of the same order of magnitude as  $\varepsilon_i$  (see p. 40).

Paying regard to (198), we get  $D_i^2 = \sum_{t=1}^{\infty} (L[a_t^{(i)}])^2$ , where

$$L[x(t)] = x(t) + b_1 \cdot x(t-1) + \dots + b_{h-1} \cdot x(t-h+1) + b_h \cdot x(t-h).$$

Write next  $a_t^{(i)}$  on the form  $\Sigma = \sum_n H^{(k_n-1)}(t) \cdot (x_n^{(i)})^t$ , where  $H^{(k)}$  is a polynomial of order  $k$ , and  $k_n$  the multiplicity of the root  $x_n$  (cf. (33)). Inserting  $a_t^{(i)}$  in  $D_i^2$ , the terms in  $\Sigma$  of type  $t^k \cdot (x_n^{(i)})^t$ , where  $x_n^{(i)} = x_n$ , will cancel out because they satisfy the relations  $L[t^k \cdot (x_n^{(i)})^t] = 0$ . It remains to estimate the terms involving  $t^k \cdot (x_n^{(i)})^t$ , where  $x_n^{(i)} = x_n - \Delta_i$ . According to the inequality of SCHWARZ it is evidently sufficient to verify that, for any of the  $k$ -values in question,  $S = \sum_{t=1}^{\infty} (L[t^k \cdot (x_n - \Delta_i)^t])^2 \rightarrow 0$  as  $i \rightarrow \infty$ .

To prove this, remove the factors  $(x_n - \Delta_i)^{t-h}$  from  $L[t^k \cdot (x_n - \Delta_i)^t]$ , and develop the remaining terms  $(x_n - \Delta_i)^s$  according to the binomial theorem. Then we get

$$S = \sum_{t=1}^{\infty} (x_n - \Delta_i)^{2t-2h} \cdot [P(t, k) + P_1(t, k) \cdot \Delta_i + \dots + P_k(t, k) \cdot \Delta_i^k]^2,$$

where

$$P(t, k) = x_n^{-t+h} \cdot L[t^k \cdot x_n^t] = x_n^h \cdot t^k + b_1 \cdot x_n^{h-1} \cdot (t-1)^k + \dots + b_h \cdot (t-h)^k.$$

Evidently  $P(t, k) = 0$  for all  $t$ . Disregarding constant factors, we have further  $P_1(t, k) = h \cdot x_n^h \cdot t^k + (h-1)b_1 \cdot x_n^{h-1} \cdot (t-1)^k + \dots + b_{h-1} \cdot x_n \cdot (t-h+1)^k$ ,  $P_2(t, k) = h(h-1) \cdot x_n^h \cdot t^k + (h-1)(h-2)b_1 \cdot x_n^{h-1} \cdot (t-1)^k + \dots + 2b_{h-2} \cdot x_n^2 \cdot (t-h+2)^k$ , etc.

Paying regard to the identities  $P(t, k) \equiv P(t, k-1) \equiv \dots \equiv P(t, 0) \equiv 0$ , we get without difficulty  $P(t+1, k+1) - P(t, k+1) = 0$  for all  $t$ . Hence  $P(t, k+1)$  must reduce to a constant, say  $c_0$ . In the same way we get  $P(t+1, k+2) - P(t, k+2) = c_1 \cdot P(t, k+1)$ , which shows that  $P(t, k+2)$  is linear in respect to  $t$ , and by induction we find that  $P(t, k+s)$  is a polynomial in  $t$  of order  $s-1$ .

Elementary transformations show that  $c_0 = P(t, k+1) - (t-h) \cdot P(t, k) = = P_1(t, k)$  for all  $t$ . Similarly,  $P(t, k+2) - (t-h+1) \cdot P_1(t, k) = P_2(t, k)$  is linear in respect of  $t$ , and repeating the procedure we find that  $P_s(t, k)$  is of order  $s-1$ .

We conclude that  $S$  is linearly composed of a finite number of terms of type  $S_k = A_i^{k+2} \cdot \sum_{t=1}^{\infty} t^k \cdot (x_n - A_i)^{2t}$ , where  $k \geq 0$ . Hence  $|S_k| \leq \varepsilon_i^{k+2} \cdot \sum_{t=1}^{\infty} t^k \cdot (1-\varepsilon_i)^{2t} \leq A \cdot \varepsilon_i^{k+2} / [1 - (1-\varepsilon_i)^2]^{k+1}$ , where  $A$  is independent of  $t$ . It is seen that for any finite  $k \geq 0$  the term  $|S_k|$  is of an order  $\leq \varepsilon_i$ , which completes the proof.

The above analysis has shown that if (251) has no root  $x_k$  falling outside the unit circle, certain well-defined linear operations on the moving average given by (249) will yield the primary process  $\{\eta(t)\}$ . If  $|x_k| \leq 1$  for all  $k$ , the sequence  $(b)$  and the process  $\{\zeta(t)\}$  will be termed *regular*.

In order to prepare the analysis of non-regular sequences  $(b)$ , let (249) be an arbitrary process of moving averages, and let its autocorrelation coefficients be represented by  $r_k$ . Paying regard to (250), we obtain the fundamental identity

$$(256) \quad \frac{1}{K^2} (x^h + b_1 x^{h-1} + \cdots + b_{h-1} x + b_h) (b_h x^h + b_{h-1} x^{h-1} + \cdots + b_1 x + 1) = \\ = r_h x^{2h} + r_{h-1} x^{2h-1} + \cdots + r_1 x^{h+1} + x^h + r_1 x^{h-1} + \cdots + r_{h-1} x + r_h.$$

Replacing here  $x$  by  $e^{ix}$ , formula (121) shows that we get  $W'(x)$ .

Until further notice, we shall again assume that (249) is regular. Denoting as before the zeros of the first factor by  $x_k$ , it is seen that the zeros of the second factor are given by  $1/x_k$ . Consequently, the zeros of the right member may be denoted  $x_1, \dots, x_{2h-1}, x_{2h}$ , where

$$x_k = 1/x_{2h-k}, \quad 0 < |x_1| \leq |x_2| \leq \cdots \leq |x_h| \leq 1 \leq |x_{h+1}| \leq \cdots \leq |x_{2h}|.$$

Further, it follows that if there exists another sequence, say  $(1, b_1^{(i)}, \dots, b_h^{(i)})$ , such that the corresponding moving average will have autocorrelation coefficients coinciding with those of (249), then one zero of the polynomial  $x^h + b_1^{(i)} x^{h-1} + \cdots + b_{h-1}^{(i)} x + b_h^{(i)}$  will equal either  $x_1$  or  $1/x_1$ , another either  $x_2$  or  $1/x_2$ , etc. Evidently, there are at most  $2^h$  real sequences of this type, say  $(b_k^{(0)}), \dots, (b_k^{(s)})$ . Letting  $(b_k^{(0)})$  represent the regular sequence started from, a short reflection reveals that all the other sequences in the group are non-regular.

If in this way a group  $(b_k^{(i)})$  of sequences is attached to every regular sequence  $(b_k^{(0)})$ , it is clear that an arbitrary sequence  $(b_k^*) =$

$(1, b_1^*, \dots, b_h^*)$  will belong to one, and only one, of these groups. It should further be observed that this group may be constructed with the use of only the sequence  $(b_k^*)$ . Similarly, a group  $(b_k^{(i)})$  will evidently be uniquely determined by the corresponding sequence of autocorrelation coefficients.

Thus prepared, let  $\{\eta(t)\}$  be a purely random process, and let  $(b_k^{(i)})$  be a group of finite sequences as defined above. Writing

$$(257) \quad [K^{(i)}]^2 = 1 + [b_1^{(i)}]^2 + [b_2^{(i)}]^2 + \dots + [b_h^{(i)}]^2,$$

let, correspondingly, a group  $(\zeta^{(i)}) = (\{\zeta^{(i)}(t; \eta)\})$  of moving averages be defined by

$$(258) \quad \zeta^{(i)}(t; \eta) = \frac{K^{(i)}}{K^{(i)}} [\eta(t) + b_1^{(i)} \eta(t-1) + \dots + b_h^{(i)} \eta(t-h)].$$

Marking the symbols referring to different processes in a group by corresponding indices, it follows from the construction of the group  $(\zeta^{(i)})$  that

$$(259) \quad D(\zeta^{(i)}) = D(\zeta^{(j)}); \quad r_k^{(i)} = r_k^{(j)}, \quad k = 0, \pm 1, \pm 2, \dots$$

Further, the group will contain one, and only one regular process, viz.  $\{\zeta^{(0)}(t)\}$ . This will be assumed to be given by (249), and be alternatively denoted by  $\{\zeta(t)\}$ .

If all roots of (251) are falling on the periphery of the unit circle, the group  $(\zeta^{(i)})$  will evidently contain only the process  $\{\zeta^{(0)}(t)\} = \{\zeta(t)\}$ . Otherwise the group will include more than one process, at most  $2^h$  in number. Furthermore, it should be observed that an equivalent construction of the group  $(\zeta^{(i)})$  is possible on the basis of the primary process  $\{\eta(t)\}$  and the characteristics (259) common to the processes in the group.

Referring to the autoregression analysis as set forth in section 19, it is seen that the coefficients in the formulae for the residuals  $\eta(t; n)$  involve only the autocorrelation coefficients and the dispersion of the process under analysis. Let this fact be combined with the above observation that the limit residual  $\lim_{n \rightarrow \infty} \eta(t; n)$  of a regular process of moving averages may be obtained directly, viz. either from (200) or from (255). We conclude that the autoregression residuals, say  $\{\eta^{(i)}(t)\}$ , of the non-regular processes in a group  $(\{\zeta^{(i)}(t; \eta)\})$  will be given by corresponding linear operations, and that these

expressions will involve exactly the same coefficients ( $a$ ) as in the case of the regular average.

Considering in the first place the case when all the roots of the equation (251) obtained from the regular process (249) in the group are falling in the interior of the unit circle, the above argument yields

$$(260) \frac{K^{(i)}}{K^{(0)}} \cdot \eta^{(i)}(t) = \eta(t) + (a_1 + b_1^{(i)}) \cdot \eta(t-1) + (a_2 + a_1 b_1^{(i)} + b_2^{(i)}) \cdot \eta(t-2) + \dots + \\ + (a_h + a_{h-1} b_1^{(i)} + \dots + b_h^{(i)}) \cdot \eta(t-h) + \\ + (a_{h+1} + a_h b_1^{(i)} + \dots + a_1 b_h^{(i)}) \cdot \eta(t-h-1) + \dots$$

Since  $(b^{(i)}) \neq (b)$ , the residual  $\eta^{(i)}(t)$  cannot reduce to  $\eta(t)$ . Further, were  $\eta^{(i)}(t)$  a finite moving average, it would follow that  $r_k(\eta^{(i)}) = 0$  for some  $k > 0$ . As  $\eta^{(i)}$  is non-autocorrelated this is impossible, so  $\eta^{(i)}$  as given by (260) must be an infinite moving average (cf. section 15 γ). Now, paying regard to the relations (250) and (252—254), an elementary transformation will verify that  $r_k(\eta^{(i)}) = 0$  for  $k \neq 0$ .

On the other hand, if at least one root of (251) is lying on the periphery of the unit circle, the representation (255) gives

$$\{\eta^{(i)}(t)\} = \lim_{k \rightarrow \infty} [\{\zeta^{(i)}(t)\} + a_1^{(k)} \{\zeta^{(i)}(t-1)\} + a_2^{(k)} \{\zeta^{(i)}(t-2)\} + \dots].$$

By construction, the coefficients  $B_n^{(k)}$  connected with (255) are such that  $\lim_{k \rightarrow \infty} B_n^{(k)} = b_n^{(0)} = b_n$ . Keeping this in mind, we conclude from (96) and (97) that  $\lim_{k \rightarrow \infty} a_n^{(k)} = a_n$ . Now, the relation (255) implies that

we may express the  $\zeta$ 's in terms of the  $\eta$ 's, and that the resulting sum, say  $\sum \{\eta(t-k)\} \cdot c_k^{(i)}$ , has a limit equalling the sum of the limits  $\{\eta(t-k)\} \cdot \lim c_k^{(i)}$ . Since  $\{\zeta^{(i)}\}$  and  $\{\zeta\}$  have identical autocorrelation properties, we may in the above relation perform the same procedure on  $\{\zeta^{(i)}\}$ . It follows that the representation (260) holds even in this case, without a limit passage being required.

Having now illustrated theorem 6 by means of a process of moving averages, the representation secured by theorem 7 is readily obtained directly. In fact, since the coefficients ( $b$ ) in the canonical formula in theorem 7 are derived solely from the dispersion and the autocorrelation coefficients of the process considered, we conclude that they must be identical for all processes in a group  $(\zeta^{(i)})$  as defined above. Thus we have (cf. also (258))

$$(261) \quad \{\zeta^{(i)}(t)\} = \{\eta^{(i)}(t)\} + b_1 \{\eta^{(i)}(t-1)\} + \cdots + b_h \{\eta^{(i)}(t-h)\}$$

for all processes in the group  $\{\zeta^{(i)}\}$ . It should be noticed that (254) gives the coefficients  $(b)$  in terms of the sequences  $(a)$  and  $(r)$  characteristic to the group.

Summing up the main results, the analysis gives the following answers to the questions set forth on p. 123. The dispersion and the autocorrelation coefficients of a process of moving averages being given, there will in general exist a well-defined group of moving averages with the same characteristics, and with the same primary process. These moving averages are limited in number, and it is possible to construct the corresponding sequences of coefficients  $(b)$  by means of the characteristics prescribed. Alternatively, if one sequence  $(b)$  in a group is known, the others are uniquely determined. — Among the moving averages in a group it is possible to distinguish one, the regular process, which alone has the property that the primary process will be given either by a relation (200) or by a limit relation of the same type. The coefficients  $(a)$  in these representations are uniquely determined by the coefficients  $(b)$  of the regular process. — For the non-regular processes in a group, there exists no relation of type (200) yielding the primary process. However, inserting a non-regular process  $\{\zeta^{(i)}(t)\}$  in the representation of the primary process in terms of the regular process belonging to the same group, we get the non-autocorrelated residual of  $\{\zeta^{(i)}(t)\}$  secured by theorem 6. According to theorem 7, the process  $\{\zeta^{(i)}(t)\}$  may be looked upon as a moving average of its residual. The coefficients of this average are nothing else than the coefficients  $(b)$  of the regular process in the group around  $\{\zeta^{(i)}(t)\}$ .

In connexion with the applications in section 31, we shall derive a linear relation of more general type than (200), which yields the primary process in terms of a non-regular moving average.

The autoregression analysis has given us no tool for distinguishing between the different moving averages in a group. If more precise information is required, other methods have to be applied, e.g. an analysis of conditioned variables (cf. p. 164). Such lines of research falling outside of the scope of the present study, this section will be terminated by some explicit examples of the previous analysis.

*Illustrations.* 1) Let the autocorrelation coefficients of a moving average be given by  $r_1 = -5$ ,  $r_n = 0$  for  $n > 1$ .

The relation (256) reads in this case

$$\cdot 5(x-1)(-x+1) = -\cdot 5x^2 + x - \cdot 5.$$

We conclude that the group  $(b^{(i)})$  contains only one sequence, viz.  $(1, -1)$ . Thus, the general formula of type (249) for a moving average with the given autocorrelation coefficients reads

$$(262) \quad \{\zeta(t)\} = \{\eta(t)\} - \{\eta(t-1)\}.$$

The equation (251) has only one root,  $x_1=1$ , and this falls on the unit circle. Hence, in order to express  $\eta$  in terms of  $\zeta$ , we have to apply the limit relation (255). Taking, for instance,  $x_1^{(k)}=1-10^{-k}$ , we get

$$(263) \quad \{\eta(t)\} = \lim_{k \rightarrow \infty} [\{\zeta(t)\} + (1-10^{-k}) \cdot \{\zeta(t-1)\} + (1-10^{-k})^2 \cdot \{\zeta(t-2)\} + \dots].$$

Table 2 contains a model series section of a process of type (262). Taking  $k=1$ , and applying (263) to the last element but one in this section, it is seen that we get the following approximation to the last element in table 1, (2),

$$-2 + 2(\cdot 9) - (\cdot 9)^3 + (\cdot 9)^4 - (\cdot 9)^5 - (\cdot 9)^6 + (\cdot 9)^7 + (\cdot 9)^8 + \dots$$

A computation of this sum has given  $-\cdot 97$ , which is fairly close to the exact value, i. e.  $-1$ . In point of principle, it is possible to reconstruct in this way the model series  $(\alpha_t^{(2)})$  on the basis of the model series  $(\beta_t)$ .

2) Let the coefficients  $(b)$  of a moving average (249) be given by (1, 2).

Forming the relation (256), we get

$$\cdot 2(x+2)(2x+1) = \cdot 4x^2 + x + \cdot 4,$$

and conclude that  $r_1=\cdot 4$ ,  $r_n=0$  for  $n>1$ . The group  $(b^{(i)})$  is seen to consist of  $(1, \cdot 5)$  and  $(1, 2)$ , the former sequence being the regular one. Now, the system (97) gives  $a_1=-\cdot 5$ , while the relations corresponding to (96) show that  $a_n=(-\cdot 5)^n$ . Considering the general formula for the regular process,

$$\{\zeta(t)\} = \{\eta(t)\} + \cdot 5\{\eta(t-1)\},$$

it follows that

$$\{\eta(t)\} = \{\zeta(t)\} - \cdot 5\{\zeta(t-1)\} + (\cdot 5)^2\{\zeta(t-2)\} - (\cdot 5)^3\{\zeta(t-3)\} + \dots,$$

in full agreement with (200). Observing that  $K^2=1\cdot 25$ , and that  $(K^{(1)})^2=5$ , formula (258) gives for the remaining process in the group  $(\zeta^{(i)})$

$$(264) \quad \{\zeta^{(1)}(t)\} = \cdot 5\{\eta(t)\} + \{\eta(t-1)\}.$$

According to the general analysis, the residual  $\{\eta^{(1)}(t)\}$  secured by theorem 6 is obtained by replacing  $\zeta$  by  $\zeta^{(1)}$  in (200). As is readily verified, the corresponding representation of type (260) reads

$$\eta^{(1)}(t) = \frac{1}{2}\eta(t) + \frac{3}{2}\eta(t-1) - \frac{3}{4}\eta(t-2) + \frac{3}{8}\eta(t-3) - \dots$$

We obtain, in analogy with (264), and in harmony with the general formula (261)

$$\{\zeta^{(1)}(t)\} = \{\eta^{(1)}(t)\} + \cdot 5 \{\eta^{(1)}(t-1)\} = \cdot 5 \{\eta(t)\} + \{\eta(t-1)\}.$$

It can be easily verified that  $D^2(\eta^{(1)}(t)) = D^2(\eta(t))$ , and that the process  $\{\eta^{(1)}(t)\}$  is non-autocorrelated, i. e. that  $r_k(\eta^{(1)}) = 0$  for  $k \neq 0$ .

3) Next, let  $r_1 = \frac{1}{6}$ ;  $r_2 = -\frac{1}{3}$ ;  $r_k = 0$  for  $k > 2$ .

The relation (256) reads in this case

$$(x^2 + \cdot 5x - \cdot 5)(-\cdot 5x^2 + \cdot 5x + 1) = -\frac{1}{3}x^4 + \frac{1}{6}x^3 + x^2 + \frac{1}{6}x - \frac{1}{3}.$$

Paying regard to the identity  $x^2 + \cdot 5x - \cdot 5 = (x - \cdot 5)(x + 1)$ , a short calculation will show that the group  $(b^{(i)})$  consists of two sequences, viz.  $(1, \cdot 5, -\cdot 5)$ , which is the regular one, and  $(1, -1, -2)$ . The general formula for a corresponding regular process reads

$$\{\zeta(t)\} = \{\eta(t)\} + \cdot 5 \{\eta(t-1)\} - \cdot 5 \{\eta(t-2)\}.$$

We find without difficulty

$$a_k = \frac{1}{3}(\frac{1}{2})^k + \frac{2}{3}(-1)^k.$$

The non-regular process is given by

$$\begin{aligned} \{\zeta^{(1)}(t)\} &= \cdot 5 \{\eta(t)\} - \cdot 5 \{\eta(t-1)\} - \{\eta(t-2)\} = \\ &= \{\eta^{(1)}(t)\} + \cdot 5 \{\eta^{(1)}(t-1)\} - \cdot 5 \{\eta^{(1)}(t-2)\}. \end{aligned}$$

Formula (260) gives for the non-autocorrelated residual

$$\eta^{(1)}(t) = \frac{1}{2}\eta(t) - \frac{3}{4}\eta(t-1) - \frac{8}{3}\eta(t-2) - \frac{3}{16}\eta(t-3) - \dots$$

4) A few remarks in connexion with the illustrations concluding Chapter II will be sufficient to show how the general formulae given in the present chapter will work in the case of a normal process of moving averages. The developments below cover both processes of linear autoregression and processes of moving averages.

The matrix of the infinite quadratic form appearing in the characteristic function of a non-autocorrelated normal process  $\{\eta\}$  with dispersion 1 is nothing else than the unit matrix. Now, keeping in mind the substitution procedure indicated on p. 90, the relations (210) will verify the product formula already given on p. 91,

$$\begin{pmatrix} 1 & b_1 & b_2 & \dots \\ 0 & 1 & b_1 & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \cdot \begin{pmatrix} x^2 & 0 & 0 & \dots \\ 0 & x^2 & 0 & \dots \\ 0 & 0 & x^2 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & \dots \\ b_1 & 1 & 0 & \dots \\ b_2 & b_1 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} = \begin{pmatrix} 1 & r_1 & r_2 & \dots \\ r_1 & 1 & r_1 & \dots \\ r_2 & r_1 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix},$$

which yields the matrix belonging to a normal process of finite or infinite moving averages. The previous illustrations exemplify the fact that there are, in general, several sequences  $(b)$  giving rise to the same matrix in the right member.

The general inversion formula (200) implies that

$$(265) \quad \begin{pmatrix} 1 & a_1 & a_2 & \dots \\ 0 & 1 & a_1 & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \cdot \begin{pmatrix} 1 & r_1 & r_2 & \dots \\ r_1 & 1 & r_1 & \dots \\ r_2 & r_1 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & \dots \\ a_1 & 1 & 0 & \dots \\ a_2 & a_1 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} = \begin{pmatrix} x^2 & 0 & 0 & \dots \\ 0 & x^2 & 0 & \dots \\ 0 & 0 & x^2 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

The verification is immediate. In fact, paying regard to (208) and (210) we obtain in the first place

$$\begin{pmatrix} 1 & r_1 & r_2 & \dots \\ r_1 & 1 & r_1 & \dots \\ r_2 & r_1 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & \dots \\ a_1 & 1 & 0 & \dots \\ a_2 & a_1 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} = \begin{pmatrix} x^2, & x^2 \cdot b_1, & x^2 \cdot b_2, & \dots \\ 0, & x^2, & x^2 \cdot b_1, & \dots \\ 0, & 0, & x^2, & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}.$$

Multiplying the right member with the first factor in (265), and keeping in mind the relations (201), we arrive at the right member of (265).

- 5) Let  $p_i$  be the multiplicity of a real root or of a conjugate complex root-pair of (251). Then the number of averages round (249) is  $\prod_i (p_i + 1)$ , the index  $i$  running over those roots and root-pairs which are  $\neq 1$  in modulus.

## CHAPTER IV.

### On the application of some stationary schemes.

#### 27. Preliminary remarks. Disposition.

In applied time series analysis, the chief problem is to find a hypothetical scheme which from a theoretical viewpoint is appropriate to the phenomenon considered, and gives a satisfactory fit to the observational data. Another desideratum is, of course, that the hypothesis be as simple as possible.

As shown in detail in sections 15 and 16, the general stationary process embraces all of the hypotheses about time series surveyed in Chapter I. The wide scope of the stationary process is due to the fact that the restrictions are reduced to a minimum: Besides the indispensable postulate that the probability laws must not contradict themselves (see (53)–(54)), the only further assumption is that time itself will not influence these probability laws (see (55)). In other words, time is thought of as a passive medium; roughly speaking, this means that any prognosis based upon the past development will depend only on this same development — i.e. supposing that the same development had taken place with a constant lag, the corresponding forecast would not differ, apart from the displacement in time.

It stands to reason that the assumption of stationarity is legitimate in the most varied fields of scientific research. Restricting ourselves in such cases to equidistant time points, we have at our disposal all of the schemes falling under the discrete stationary process, and this is a larger class than the continuous stationary process (cf. p. 70 f).

From the viewpoint of the theory of probabilities, the purely random process is the simplest type case of a stationary scheme. In sections 15 and 16, certain other type cases were constructed

on the basis of the purely random process, and the scheme of hidden periodicities was interpreted as a stationary process. The special theoretical models mentioned are rather simple, in as much as an adequate description of their structure is possible by means of linear methods, such as the periodogram analysis and the analysis of the graph of autocorrelation coefficients.

The present chapter is reserved for some applications of the above-mentioned simple schemes, in particular the schemes of linear regression. It will be seen that the results obtained are rather promising. We advance also that certain points in the applications will give rise to theoretical discussions — the analysis in the previous chapters was chiefly concerned with the structural properties of the hypothetical models, while questions bearing upon their application were touched upon only incidentally.

Of course, in applying different hypothetical schemes to observational data, different methods are required.

In the search for a hypothetical model suitable to a stationary phenomenon, the construction of an empirical periodogram is a classical method of fundamental importance. A careful periodogram construction is a safe method for discovering hidden periodicities if such are really present. On the other hand, approximate methods often involve definite dangers. For instance, the BRUNS-OPPENHEIM method (see section 5) fails as often as the periodic elements are covered by a random component. The bias in question, which was already observed by J. I. CRAIG (1916), is of interest also in view of other methods of analysis. For this reason, the »CRAIG effect» will be examined in some detail. This is done in the next section.

As shown in detail in Chapter III, periodogram analysis is an inadequate method of research in the cases of linear autoregression and of moving averages. The present study being focussed on these schemes, we have to look for appropriate substitutes for the periodogram method. In the memoir where G. U. YULE (1927) introduces the scheme of linear autoregression, an empirical parallel to the autoregression analysis as developed in section 19 forms the leading method of research. This method yields a first substitute for the periodogram construction. Next, as emphasized by SIR G. WALKER (1931), the autocorrelation coefficients behave quite differently in the schemes of hidden periodicities and linear autoregression. Hence, the graph of the serial coefficients will yield important information

about the nature of the phenomenon considered. The analysis in section 26 shows that the scheme of moving averages, too, presents a characteristic graph of autocorrelation coefficients, a circumstance which augments the importance of the method proposed by WALKER. For the sake of brevity in writing, the graphs of serial and autocorrelation coefficients will be termed *correlograms* (*empirical* and *hypothetical* respectively).

In the following, the methods proposed by G. U. YULE (1927) and SIR G. WALKER (1931) will be carried further on the basis of the theoretical investigations in the previous chapters, and used in the applications to empirical data. A critical survey of the original methods of YULE and WALKER will be given in section 29. In section 30 follow a summary and a critical examination of the modified methods.

The two sections concluding the present study are reserved for applications of the scheme of moving averages and the scheme of linear autoregression. Both these schemes of linear regression may be attached directly to familiar lines of time series analysis. Recurring to this point later, especial attention will be drawn to the different type of forecast yielded by these schemes as compared with the scheme of hidden periodicities. While the forecasts delivered by the latter scheme cover an infinite future, the former schemes will yield efficient forecasts only over a limited period of time. However, this limitation is outweighed by the greater efficiency in the short time forecasts yielded by the schemes of linear regression.

Finally, in discussing the applications, certain generalizations of the schemes of linear regression will be touched upon.

## 28. On the CRAIG effect.

In section 5 we have surveyed a few methods for separating the individual components in a sum of harmonics (17). The classical method being the construction of a periodogram, the short cut indicated by S. OPPENHEIM (1909) (see p. 18) is based on the difference relations satisfied by the function (17).

The scheme of hidden periodicities consists of a composed harmonic on which a purely random component is additively superposed.

Even in this case a periodogram will point out the frequencies of the individual harmonics. But as emphasized by J. I. CRAIG (1916), the OPPENHEIM method will be biased by the random component. Proceeding to an examination of this bias, which will be termed the »CRAIG effect», it will be sufficient for our purpose to consider a scheme of simple structure. In doing this, we shall regard the composed harmonic as a sample series of a singular process. This device, which will not affect the proof, is used merely in order to illustrate the connexion between the scheme of hidden periodicities as defined in section 8 and the process of hidden periodicities as defined in section 15 Ω.

Let  $\{\psi(t)\}$  stand for a singular process satisfying the relation

$$\mathcal{A}^2 \psi(t-1) + k^2 \cdot \psi(t) = 0, \quad 0 < k \leq 2.$$

Let  $\{\eta(t)\}$  be purely random, and independent of  $\{\psi(t)\}$ . Let further,  $\{\eta(t)\}$  have a finite dispersion and a vanishing mean, and let a process of hidden periodicities be defined by

$$(266) \quad \{\xi(t)\} = \{\psi(t)\} + \{\eta(t)\}.$$

Denoting by  $\psi_i(t, t-1, \dots, t-n) = [\psi_i(t), \psi_i(t-1), \dots, \psi_i(t-n)]$  a sample series section of the process  $\{\psi(t)\}$ , we have

$$\mathcal{A}^2 \psi_i(t-s-1) + k^2 \psi_i(t-s) = E[\mathcal{A}^2 \psi(t-1) + k^2 \psi(t)] = 0.$$

Now, let  $k^2$  be unknown. The value delivered by the OPPENHEIM method is that minimizing the expression

$$\frac{1}{n-1} \sum_{s=1}^{n-1} [\mathcal{A}^2 \psi_i(t-s-1) + k^2 \psi_i(t-s)]^2.$$

A short reduction leads to the following value which is obviously unbiased,

$$(267) \quad k^2 = - \sum_{s=1}^{n-1} \psi_i(t-s) \cdot \mathcal{A}^2 \psi_i(t-s-1) / \sum_{s=1}^{n-1} \psi_i^2(t-s) = \\ = - E[\psi(t) \cdot \mathcal{A}^2 \psi(t-1)] / E[\psi^2(t)].$$

According to (31), the frequency  $\lambda$  of the harmonic  $\psi_i(t)$  is given by

$$(268) \quad \lambda = 2 \arcsin k/2.$$

Let it next be assumed that a sample series  $\xi_i(t, t-1, \dots, t-n)$  is given, and that we know neither  $k^2$  nor the sample series  $\psi_i$  and  $\eta_i$ . The assumption concerning an additive random element corresponds to the actual situation when seeking for a period in an empirical time series. Applying now the OPPENHEIM method, we must minimize

$$(269) \quad \frac{1}{n-1} \cdot \sum_{s=1}^{n-1} [\mathcal{A}^2 \xi_i(t-s-1) + \bar{k}^2 \cdot \xi_i(t-s)]^2,$$

which gives (cf. (267))

$$(270) \quad \bar{k}^2 = - \sum_{s=1}^{n-1} \xi_i(t-s) \cdot \mathcal{A}^2 \xi_i(t-s-1) / \sum_{s=1}^{n-1} \xi_i^2(t-s).$$

This expression depends on the actual path of the sample series  $\xi_i(t, t-1, \dots, t-n)$ , but a sufficient approximation is delivered by the  $\bar{k}^2$ -value minimizing the expression

$$(271) \quad E_i [\mathcal{A}^2 \psi(t-1) + \bar{k}^2 \psi(t)]^2 + E [\mathcal{A}^2 \eta(t-1) + \bar{k}^2 \eta(t)]^2,$$

where we have written  $E_i [f[\psi]]$  for  $\lim_{n \rightarrow \infty} \frac{1}{n-1} \cdot \sum_{s=1}^{n-1} f[\psi_i(t-s)]$ .

A short calculation shows that this approximation reads

$$(272) \quad \bar{k}^2 \sim [-E_i \psi(t) \cdot \mathcal{A}^2 \psi(t-1) + 2 E \eta^2(t)] / [E_i \psi^2(t) + E \eta^2(t)].$$

The difference between on one hand (270) and (272) and on the other hand the unbiased formula (267) gives rise to the CRAIG effect. Formula (272) shows that this effect will depend on the chance-determined amplitude of the harmonic constituting  $\psi_i(t, t-1, \dots, t-n)$ .

It is seen that the  $\bar{k}^2$ -value given by (272) equals  $k^2$  when, and only when,  $k^2 = 2$ , i. e. when  $\psi(t)$  has the period 4 time units. A brief reflection shows further that the OPPENHEIM method will under-estimate the period if this is above 4 time units, while the reverse will be true if the period is below 4 time units. Moreover, the larger the variance of the random component  $\eta$ , the larger is the CRAIG effect.

*Illustrations.* Two simple examples of the CRAIG effect will be given on the basis of the model series  $(\bar{\Omega}_t^{(1)})$  and  $(\bar{\Omega}_t^{(2)})$  presented in table 4. Applying the relations (270), (268) and (18), the following results were obtained, the sums running from  $t=2$  to  $t=999$ .

	Model series ( $\bar{\Omega}_t^{(1)}$ )	Model series ( $\bar{\Omega}_t^{(2)}$ )
$\sum \bar{\Omega}_t \cdot \mathcal{A}^2 \bar{\Omega}_{t-1}$	- 4041	- 5568
$\sum [\bar{\Omega}_t]^2$	1112	1711
$\bar{k}^2$	3.6340	3.2542
$\bar{\lambda}$	144.8°	128.8°
$\bar{p}$	2.486	2.795

The correct value of the period being 2 units in both the time series, the CRAIG effect is seen to be substantial. The values  $\bar{k}^2$  obtained are in good agreement with the approximate formula (272). In fact, observing that in each of the two cases  $E_i[\psi(t) \cdot \mathcal{A}^2 \psi(t-1)] = -4$ , and that the variances of the random components are respectively 2 and 6, we get in the first case  $\bar{k}^2 \sim 4 \cdot 4 / 1 \cdot 2 = 3.667$ , and in the second  $\bar{k}^2 \sim 5.2 / 1.6 = 3.25$ . In full agreement with the remarks attached to formula (272), the period is over-estimated, and the CRAIG effect is larger in the second model series than in the first.

In the above analysis of the scheme (266) we have examined the OPPENHEIM method for determining the parameter  $k$  in an approach of type  $\mathcal{A}^2 \xi(t-1) \sim k \cdot \xi(t)$  or  $\xi(t) \sim (2-k) \cdot \xi(t-1) - \xi(t-2)$ . We shall next apply the same method in starting from a more general approach, viz.

$$(273) \quad \xi(t) \sim a_1 \cdot \xi(t-1) + a_2 \cdot \xi(t-2).$$

Letting as before  $\xi_i(t, t-1, \dots, t-n)$  represent a sample series section connected with the scheme (266) of hidden periodicities, we must in the present case minimize (cf. (269))

$$(274) \quad \frac{1}{n-1} \sum_{s=0}^{n-2} [\xi_i(t-s) - \bar{a}_1 \cdot \xi_i(t-s-1) - \bar{a}_2 \cdot \xi_i(t-s-2)]^2$$

in respect of  $\bar{a}_1$  and  $\bar{a}_2$ . Now, using an approximation of the same type as in (272), and paying regard to (46), we get

$$(275) \quad r_1 - \bar{a}_1 - \bar{a}_2 \cdot r_1 \sim 0, \quad r_2 - \bar{a}_1 \cdot r_1 - \bar{a}_2 \sim 0,$$

having written  $r_k = \cos \lambda k / (1 + d_i^2)$ . Here  $\lambda$  is the frequency of a sample series connected with the singular process  $\{\psi(t)\}$ , and  $d_i^2 = D^2(\eta)/D_i^2(\psi)$ , where  $D_i(\psi)$  is the dispersion in the sample series  $\psi_i(t, t-1, \dots)$ . Solving (275), we obtain

$$(276) \quad \begin{aligned} \bar{a}_1 &\sim 2 \cdot \frac{\sin^2 \lambda + d_i^2}{\sin^2 \lambda + 4d_i^2 + 4d_i^4} \cdot \cos \lambda; \\ \bar{a}_2 &\sim -\frac{\sin^2 \lambda - 2d_i^2 \cdot \cos 2\lambda}{\sin^2 \lambda + 4d_i^2 + 4d_i^4} \geq -1. \end{aligned}$$

The minimizing of (274) is obviously analogous to that phase of the autoregression analysis described in section 19, where  $\xi(t)$  is linearly approximated by  $\xi(t-1)$  and  $\xi(t-2)$ . It is in view of this analogy that the above formulae are of interest. Now, if  $D(\eta) = 0$ , we obtain from (276) the coefficients  $\bar{a}_1 = 2 \cos \lambda$  and  $\bar{a}_2 = -1$  appearing in the identity (cf. (35))

$$\xi(t) = \psi(t) = 2 \cos \lambda \cdot \psi(t-1) - \psi(t-2).$$

It should further be observed that we have in this case  $\cos \lambda = -\bar{a}_1 / 2\sqrt{-\bar{a}_2}$ . On the other hand, if  $D(\eta) \neq 0$ , the latter relation will be disturbed by a CRAIG effect. In fact, we get from (276)

$$(277) \quad \cos \bar{\lambda} = \frac{(\sin^2 \lambda + d_i^2) \cdot \cos \lambda}{\sqrt{(\sin^2 \lambda - 2d_i^2 \cdot \cos 2\lambda)(\sin^2 \lambda + 4d_i^2 + 4d_i^4)}}.$$

Approximating the period  $p$  of  $\psi(t)$  by means of the biased frequency  $\bar{\lambda}$  given by (277), the CRAIG effect is seen to be particularly large if  $\sin \lambda$  is small. It would serve no purpose to discuss the sign of the deviation or to enter into details on a singular process  $\{\psi(t)\}$  of general structure.

*Illustration.* Considering the model series  $(\bar{Q}_t^{(1)})$  dealt with in the previous illustration, we have found by minimizing (274)  $-910 = \bar{a}_1 \cdot 1114 - \bar{a}_2 \cdot 911$ ;  $930 = -\bar{a}_1 \cdot 911 + \bar{a}_2 \cdot 1114$ . This system gives  $\bar{a}_1 = -4051$ ,  $\bar{a}_2 = 5036$ . Since  $a_1 = -2$ , and  $a_2 = -1$  the method examined is completely misleading. Formula (276) explains the failure — in the present case  $\sin \lambda = 0$ , and the resulting  $\bar{a}$ -values will show no tendency whatever to approximate the  $a$ -values sought for. As is readily verified, (276) gives  $-\bar{a}_1 \sim \bar{a}_2 \sim 4167$ .

Summing up the above analysis, we conclude that the OPPENHEIM method yields no adequate substitute for a periodogram construction in case the periodicities are covered by a random element, and it does not seem worth the trouble to derive modifications neutralizing the bias involved.

## 29. On earlier applications of the scheme of linear autoregression.

As far as I am aware, there are only two earlier investigations which present direct applications of the schemes of linear regression to empirical data, viz. those already referred to by G. U. YULE (1927) and SIR G. WALKER (1931). These memoirs are concerned with sunspots and air pressure respectively, and in both cases it is the scheme of autoregression that is applied. An examination of the main lines of these investigations in the light of the previous analysis will be given in the present section. For the sake of completeness, we shall also touch upon a passage in the already mentioned study on expectance theory by K. STUMPF (1936), where an empirical correlogram is dealt with by use of a method related to that proposed by SIR G. WALKER.

The memoir of G. U. YULE (1927) starts with a discussion of a model series, say  $\xi_t$ , constructed on the basis of a relation of type

$$(278) \quad \xi_t - k \xi_{t-1} + \xi_{t-2} = \eta_t, \quad -2 < k < 2,$$

the purely random series  $\eta_t$  being given by dice-throwing. The constant  $k$  is chosen in the interval  $(-2, 2)$ , which implies that the roots of the characteristic equation of (278) are complex, and of modulus unity. Thus it follows from the general analysis in section 22 that a process  $\{\xi(t)\}$  corresponding to (278) is non-stationary. However, the evolutive tendency is rather weak, and the 300 elements constituting YULE's model series actually present fluctuations of a stationary appearance.

The parameter  $k$  in the model series  $\xi_t$  being chosen so as to correspond to a period of 10 time units, YULE lays stress upon the structural resemblance between his model series and the yearly index of sunspots. Pursuing this suggestion in the later sections of his memoir, he gives two different methods for a refined analysis of the structure of the index. YULE works on the A. WOLFER index 1751—1923.

In his first approach, YULE starts from the hypothesis that the sunspot index, say  $\xi_t$ , satisfies a relation of type (278). In order to determine  $k$ , he minimizes the sum of the squared »disturbances»  $\eta_t$ . Interpreting the relation (278) as ruling the movement of a

pendulum subjected to random shocks, YULE derives the intrinsic period of the pendulum which would correspond to the  $k$ -value obtained. The period thus derived being too short, viz.  $p = 10.08$  years, he finds that the hypothesis (278) gives a better value,  $p = 11.03$  years, when applied after graduating the sunspot numbers.

Having thus assumed a linear regression (278) between  $\xi_t + \xi_{t-2}$  and  $\xi_{t-1}$ , YULE applies a graphic test of this approach. Referring to the graphs given on p. 277, he says on the same page: »On the whole, however, divergence from linearity does not look as if it would be a serious trouble».

In his second approach, YULE starts from the relation

$$(279) \quad \xi_t + a_1 \xi_{t-1} + a_2 \xi_{t-2} = \bar{\eta}_t.$$

Proceeding as in the case (278), he determines the parameters ( $a$ ) by minimizing  $\sum \bar{\eta}_t^2$ , and interprets the results by the use of the analogy with a pendulum. The values found for  $a_1$  and  $a_2$  correspond to a damped intrinsic oscillation. The ungraduated index gives the period  $p = 10.600$  years, while the graduated index as before gives a better value,  $p = 11.164$  years.

The graduated index gives rise to a smaller variance in the disturbances  $\bar{\eta}_t$  than the ungraduated index. Dividing the variance of  $\bar{\eta}_t$  by the variance of the sunspot index under analysis, the approach (278) gives .243 for the ungraduated index, and .115 for the graduated one. The corresponding values in the approach (279) are .198 and .102 respectively.

In applying generalized hypotheses of type (278) and (279), YULE finds that the introduction of more parameters does not bring on a marked decrease in the variance of the disturbances. In other words, the experiments »fail to suggest the presence of any period other than the fundamental, a conclusion entirely in accord with the work of LARMOR and YAMAGA» (p. 295).

In a summary, YULE suggests that the sunspot numbers »should be regarded as analogous to the data that would be given by observations of a disturbed periodic movement, such as that of a pendulum subjected to successive small random impulses» (p. 294). Let us discuss this hypothesis in the light of the previous analysis.

As already pointed out, the approach (278) does not correspond to a stationary process, for if the series  $\bar{\eta}_t$  were purely random, the secondary model series  $\xi_t$  would present oscillations increasing in amplitude with time, i. e. be evolutive. In view of this observation,

it is not surprising that the disturbances  $\bar{\eta}_t$ , as calculated from (278) on the basis of the  $k$ -value previously determined, form a series of non-random character. Quoting YULE, the series  $\bar{\eta}_t$  shows »a tendency for positive disturbances during the approach to the maximum of the sunspot numbers, negative during the approach to minimum» (p. 294 f.).

We have already seen that the approach (279) gives rise to a somewhat smaller variance in the disturbances  $\bar{\eta}_t$ . However, having introduced a second parameter, the slight reduction in the variance of  $\bar{\eta}_t$  does not give a sufficient reason for preferring (279) over (278). On the other hand, (279) corresponds to a proper stationary process, which is a circumstance speaking in favour of this approach.

In itself, the outcome of significant constants  $a_1$  and  $a_2$  does not imply that the approach (279) is adequate. Without further evidence, we cannot even conclude that (279) is better than the hypothesis of a strictly periodic component in the sunspot index. In fact, our analysis in the previous section has shown that an autoregression analysis will give rise to non-vanishing coefficients ( $a$ ) also in the case of hidden periodicities. It is interesting to notice that even the effect of the graduation — the increase in the period — might be explained by assuming the index to be ruled by a scheme of hidden periodicities (cf. p. 137). However, a sufficient reason for rejecting the latter hypothesis is that the deduction of a strictly periodic component actually leaves an »error» with a dispersion substantially above that of the disturbances obtained from (279). In fact, the assumption of one harmonic component in the sunspot index would explain at most 28% of the variance in the index (see e. g. K. STUMPF (1937), p. 126). On the other hand, we have seen that the approach (279), which contains only two parameters, is able to explain at least 80% of the variance in question. In this connexion it is rather interesting to notice that the periodogram of the sunspot index given by STUMPF (l. c.) bears a certain resemblance to our fig. 4 (p. 116), and thus agrees with the hypothesis of linear autoregression. (Cf. also the remarks attached to (131)).

The values found by YULE for the parameters in (279) are  $a_1 = -1.34254$ ,  $a_2 = .65504$  for the ungraduated, and  $a_1 = -1.51527$ ,  $a_2 = .80245$  for the graduated sunspot index. Using the analogy of a swinging pendulum, these values correspond to a rather heavy

damping — in the case  $a_2 = .80245$ , the amplitude of a swing would be reduced to 29 % in the duration of one period (see G. U. YULE (1927), p. 282). With this heavy damping, a purely random series of impulses would not be likely to produce such large amplitudes as in the sunspot fluctuations. In full agreement with this argument, which in section 32 will be developed as a test of the scheme of linear autoregression, the disturbances  $\bar{\eta}_t$  calculated from (279) on the basis of the sunspot index present variations of a non-random type, in as much as they »do occur just in the kind of way that would be necessary to maintain a damped vibration» (p. 286).

Summing up the above discussion, we have seen that in replacing the approach of strict periodicity by a hypothesis containing an acting random element, G. U. YULE (1927) obtains a substantially better fit to the sunspot data. In the terminology of the present study, the approach (279) as applied to the ungraduated index corresponds to a scheme of linear autoregression. On the other hand, as applied to the graduated index, it is obvious that the approach (279) proximately corresponds to the assumption that the ungraduated index is ruled by a scheme consisting in a purely random process independent of and superposed on a process of linear autoregression.

As mentioned by YULE, the disturbances  $\bar{\eta}_t$  calculated from (279) present a certain systematic variation. This non-random behaviour, which seems to be conditioned by the small value found for  $a_2$  in (279), remains unexplained by the hypothesis of linear autoregression. In view of this circumstance, it seems to me as if the sunspot index calls for further investigation. Perhaps the methods developed in section 32 would yield a scheme fitting the data better. But it is also possible that more satisfactory results would be obtained in an approach involving a non-linear function of the index. Since to pursue these suggestions falls outside the program of this survey, we shall end our discussion of the YULE memoir.

SIR G. WALKER (1931) follows up the YULE approach (279), and studies an autoregression relation of type

$$(280) \quad \bar{\xi}_t + a_1 \bar{\xi}_{t-1} + \cdots + a_h \bar{\xi}_{t-h} = \bar{\eta}_t.$$

As mentioned in section 24, WALKER finds that the autocorrelation coefficients corresponding to (280) satisfy the relations (cf. p. 104).

$$(281) \quad r_k + a_1 r_{k-1} + \cdots + a_h r_{k-h} = 0, \quad k > h.$$

Assuming that the roots of the characteristic equation (34) are different, he further gives  $r_k$  as the general solution of (32).

We are now in a position to examine WALKER's methods for applying (280) to empirical data. The basic idea is simply to compute the serial coefficients  $\bar{r}_k$ , and to determine the constants  $a_n$  so that the relations (281) will be approximately satisfied when replacing  $r_k$  by  $\bar{r}_k$ .

SIR G. WALKER works on air pressure data from *Port Darwin* 1880—1925, taking the quarter of a year for time unit. The graph of serial coefficients — in our terminology the correlogram — ranges from  $k = 1$  to  $k = 147$  quarters (p. 531). The graph shows a rapid decrease from  $\bar{r}_1 = .76$  to  $\bar{r}_6 \cong 0$ . For larger  $k$ -values, the correlogram presents fluctuations with rather small amplitudes. In fact, up to  $k = 100$  all the coefficients  $\bar{r}_k$  are less than .3 in modulus.

SIR G. WALKER finds that in the interval  $0 \leq k \leq 40$  a fairly good approximation to the correlogram is yielded by the function

$$(282) \quad r_k = .19(.96)^k \cos \pi k / 6 + .15(.98)^k + .66(.71)^k.$$

This function, which is seen to involve a damped harmonic with period  $p = 12$  quarters, satisfies the difference equation

$$r_k - 3.35 r_{k-1} + 4.43 r_{k-2} - 2.71 r_{k-3} + .64 r_{k-4} = 0.$$

Concluding inversely from (281) on (280), WALKER finally arrives at the representation

$$(283) \quad \xi_t - 3.35 \xi_{t-1} + 4.43 \xi_{t-2} - 2.71 \xi_{t-3} + .64 \xi_{t-4} = \eta_t.$$

Proceeding to an examination of WALKER's methods, we observe in the first place that an unconditioned conclusion from (281) on (280) is not permitted. In fact, the autocorrelation coefficients corresponding to the relation (280) satisfy not only the relations (281) but also (222—223), the latter relations not having been observed by WALKER. It is seen that the coefficients  $r_1, r_2, \dots, r_{h-1}$  will be uniquely determined by the system (222) in terms of the coefficients  $a_n$ . In other words, the coefficients in (280) determine not only the periods and the damping factors of the individual components in the expression (33), but also the coefficients of the

components. Thus, without anything further we cannot be sure that the autocorrelation coefficients corresponding to the approach (283) will be given by the graduating expression (282). On the contrary, the system (222) corresponding to (283) actually gives rise to autocorrelation coefficients which are substantially different from (282). Having read off the values  $r_1 = .75$ ,  $r_2 = .55$ ,  $r_3 = .35$  from the graph of (282) given by WALKER (p. 528), I have found from (222) the values  $r_1 = .93$ ,  $r_2 = .72$ ,  $r_3 = .43$  for the autocorrelation coefficients belonging to the approach (283).

In view of the oversight pointed out, it is not surprising that the relation (283) gives rise to a larger dispersion in the disturbances  $\bar{\eta}_t$  than in the air pressure data  $\bar{\zeta}_t$ . As a matter of fact, I have found  $D(\bar{\eta}) = 2.4 D(\bar{\zeta})$ , while the result  $D(\bar{\eta}) = .28 D(\bar{\zeta})$  obtained by WALKER (p. 530) is based on an incorrect use of the relation mentioned in a note in section 25 (cf. (317)). We conclude that if the approach (283) is to be applied to the air pressure data, the coefficients must be modified. Having stated this, it is rather interesting that the simple approach

$$(284) \quad \bar{\zeta}_t - .73 \bar{\zeta}_{t-1} = \bar{\eta}_t$$

gives a fairly good fit to the first few serial coefficients. In fact, according to (283) the approach (284) gives  $r_1 = .73$ ,  $r_2 = .53$ ,  $r_3 = .39$ ,  $r_4 = .28$ , while the air pressure serial coefficients given by WALKER (p. 528) read  $\bar{r}_1 = .76$ ,  $\bar{r}_2 = .56$ ,  $\bar{r}_3 = .36$ ,  $\bar{r}_4 = .18$ . A short calculation shows further that (284) gives  $D(\bar{\eta}) = .68 D(\bar{\zeta})$ .

Let us in conclusion attach a few remarks to the empirical correlogram presented by SIR G. WALKER (p. 531). As already mentioned, the serial coefficients show rather small deviations from zero in the interval  $3 < k < 40$ . On the other hand, the increase in amplitude for certain  $k$ -values  $> 40$  might be due to the successive reduction in the number of correlates. Perhaps this argument is sufficient to explain also why the fluctuations are somewhat larger in that alternative variant of a correlogram given by WALKER, where all serial coefficients are based on 77 pairs of correlates. As the fluctuations, furthermore, seem rather irregular and aperiodic — at least to my eye — it is doubtful whether it would be possible to improve sensibly the approach (284) by taking into account more distant elements  $\bar{\zeta}_{t-2}$ ,  $\bar{\zeta}_{t-3}$ , etc. In this connexion it is rather interesting to notice that according to the general analysis there exists no process of linear autoregression having (282) for auto-

correlation coefficients. Another reason for resting satisfied with the simple approach (284) is that the ordinates in the periodogram presented by SIR G. WALKER (p. 526) are all lying on about the same level — this periodogram does not like that of the sunspots suggest a scheme of linear autoregression with a tendency to periodicity (cf. p. 142). At any rate, a more detailed analysis of the air pressure data is beyond the scope of the present survey.

Using formula (127), K. STUMPF (1936) arrives at an expectance theory which generalizes the classical SCHUSTER theory. In applying his theory, STUMPF works on air pressure data (*Potsdam*  $\frac{3}{7}$  1925— $\frac{27}{6}$  1926, equidistance 1 hour), and replaces the coefficients  $r_k$  in (127) by a corresponding set of graduated serial coefficients. Claiming that the graduated values belong to a scheme of linear autoregression of type

$$(285) \quad \zeta(t) - 2p \cdot \zeta(t-1) + p^2 \cdot \zeta(t-2) = \eta(t),$$

STUMPF makes a mistake similar to WALKER's pointed out above. Proceeding as in the developments on p. 112, I have found the formula  $r_k = [1 + k \cdot (1 - p^2)/(1 + p^2)] \cdot p^k$  for the autocorrelation coefficients belonging to the scheme (285), while K. STUMPF (p. 53) gives  $r_k = (1 - k \cdot \log p) \cdot p^k$ .

### 30. Preliminary survey of methods.

In this section we shall give a brief summary of the methods used in the later applications and a few critical remarks on the scope of these methods.

As pointed out in earlier sections, a careful analysis of the structural properties of a time series requires statistical data covering a rather long period. On the other hand, the series must not change its general character in the course of the interval of observation, for then a stationary scheme would be inadequate. For instance, if a trend is present in the material, it should be removed before starting the analysis (cf. p. 1).

Considering a scheme (39) of hidden periodicities, and disregarding the value  $r_0 = 1$ , the correlogram  $r_k$  consists of superposed harmonics such that the periods of the individual components equal those in the time series considered (cf. (46)). In the two type cases of linear

regression, on the other hand, the correlogram has the horizontal axis for asymptote. In fact, in the scheme of linear autoregression the correlogram  $r_k$  is a function (33) such that the roots of the characteristic equation (34) are of modulus less than unity, and in the scheme of moving averages all the autocorrelation coefficients are zero beyond a certain  $k$ -value (cf. also (177)).

For the sake of concreteness, we show below three hypothetical correlograms exemplifying the type cases considered.

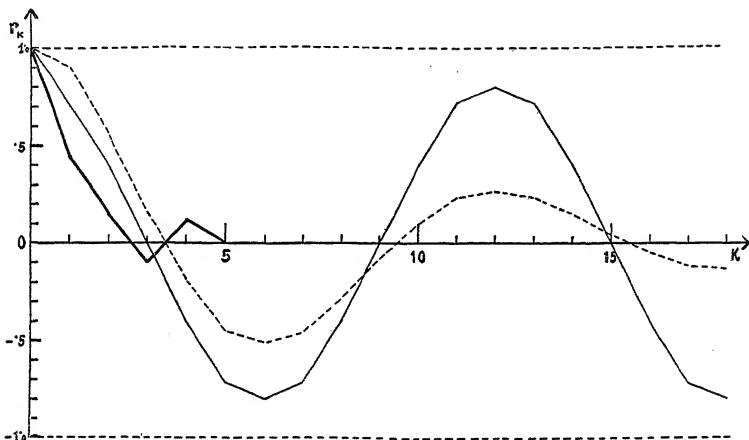


Fig. 6. Correlograms illustrating the schemes of hidden periodicities (thin line), linear autoregression (broken line), and moving averages (thick line).

The correlograms in this figure are based on the following parameters. In the case of hidden periodicities, the correlogram has been derived from formula (46), where we have chosen  $s = C_1 = 1$ ;  $D^2(\eta) = 125$ ;  $\lambda_1 = \pi/6$ . The case of linear autoregression is illustrated by a correlogram of type (243), having taken  $C = \sqrt{8}$ ,  $\lambda_1 = \pi/6$ . The moving average correlogram, finally, has been obtained from (250), putting  $h = 4$ ;  $b_1 = .7$ ,  $b_2 = .4$ ,  $b_3 = -.3$ ,  $b_4 = .2$ .

Because of the different behaviour of the autocorrelation coefficients in the schemes mentioned, it may be expected that we would obtain useful suggestions by inspecting the empirical correlogram when searching for an adequate scheme to be applied to an observational time series. For this reason, the construction of an empirical correlogram is taken as the starting point in the following applications.

It should be observed that the correlogram construction by formula (13) involves a relatively small amount of numerical computa-

tion. No trigonometrical or other mathematical tables are required. Another definite advantage is that the correlogram is obtained directly from the statistical data, without any preceding preparation of the material. Accordingly, the empirical correlogram seems particularly well suited for serving as a first indicator of which type of scheme to apply to the data.

If the empirical correlogram suggests a scheme of hidden periodicities, the next step would be to construct a periodogram for a more detailed analysis of possible periodicities in the material under investigation.

Next, if the correlogram suggests a scheme of linear autoregression, our first problem is to find a scheme (101) such that the corresponding hypothetical correlogram will fit the empirical one. The chief difficulty is to derive suitable values for the coefficients ( $\alpha$ ) — when having arrived at a set of coefficients ( $\alpha$ ), the corresponding autocorrelation coefficients will be uniquely determined by the system (221—2), and the residuals  $\eta_t$  by the relations (280). It is further a desideratum that these residuals be as small as possible. Having seen above that these problems are more intricate than emphasized in earlier studies of the graph of serial coefficients, it will be found that an empirical autoregression analysis as proposed by G. U. YULE (1927) will be useful in this connexion.

Finally, it may happen that the empirical correlogram will suggest a scheme of moving averages. As far as I know, the problem of fitting this scheme to observational data has not been attacked in earlier literature. A fundamental problem in this sphere was formulated by Prof. H. CRAMÉR in his 1933 Course, viz. to find a moving average with a prescribed correlogram. It will be seen that the relation (256) gives a starting point for attacking this problem.

Having now given a preliminary survey, the details of the methods will be discussed when presenting the results of their application. The present section will be concluded by attaching some remarks of general scope concerning the limitation of the methods outlined.

In time series analysis, significance problems are extremely intricate. In dealing with serial coefficients, we have i. a. to pay regard to the fact that their magnitude is conditioned by the size of the statistical masses to which they refer. Having already touched upon this circumstance in section 24, the point in question will be taken up for a more detailed examination in appendix B.

The following applications aim only at illustrating the qualitative

differences between various hypothetical models. Consequently, all questions about the significance and the interpretation of the quantitative results fall outside the scope of this study, and again an explicit warning is given against attaching importance to the numerical values found for the parameters of the different models fitted to the observational data.

However, even when limiting the program to an analysis of the qualitative structure of a phenomenon as described by a time series, we must be cautious when interpreting the results. Stating the case briefly: The results will to some extent be conditioned by the methods used in the analysis. Of course, in point of principle the situation is the same in all applications of hypothetical models to empirical data. Let us dwell a moment on some circumstances which are peculiar to time series analysis, especially as based on the correlogram and the autoregression methods.

The correlogram sums up the autocorrelation properties of a time series, and the autoregression analysis, too, is based solely on the autocorrelation coefficients. We conclude that neither method is able to distinguish between different schemes with coincident autocorrelation coefficients. Having already seen examples of this when dealing with the scheme of moving averages (cf. section 26), further examples are readily obtained by using non-linear operations in the construction of stationary processes. For instance, letting  $\{\eta(t)\}$  represent a purely random process, it is evident that

$$(286) \quad \xi(t) = \eta(t) \cdot \eta(t - 1)$$

will define a stationary process  $\{\xi(t)\}$ . Assuming that  $E[\eta(t)] = 0$ , a short calculation will show that the process  $\{\xi(t)\}$  is non-autocorrelated.

Thus, having found a hypothetical scheme yielding a good fit to an empirical correlogram, it is perfectly possible that there are other schemes which yield an equally close approximation. When it is necessary to choose between different schemes, it may happen that theoretical arguments will speak in favour of one of the schemes. As exemplified in the applications, the schemes of linear regression often seem plausible from theoretical viewpoints, at least as a first approximation. On the other hand, a rational choice between different schemes may be alternatively based on an examination of other structural properties of the time series than its

serial coefficients. Such lines of research, however, fall outside of the program of the present study.

Another point which must be kept in mind has special reference to the autoregression analysis. Having in sections 19 and 20 subjected a general stationary process to an autoregression analysis, the investigation resulted in a canonical form for the process considered, viz. a decomposition in two mutually non-correlated processes, each of a structure readily comprehended in respect to certain fundamental properties. Now, even if a complete parallel to this analysis could be carried through when dealing with empirical data — which of course is impossible, one reason being the necessity of dealing with only a finite number of observations — it is not certain that the autoregression analysis would be an adequate method of research. In fact, in point of principle an autoregression analysis can reveal only linear interrelations between the elements in a time series. For instance, the implicit relation

$$(287) \quad \xi(t) = \eta(t) + p \cdot \xi^2(t-1),$$

where  $\{\eta(t)\}$  is purely random,  $P[|\eta(t)| < 1] = 1$ , and  $|p| < 1$ , defines a stationary process  $\{\xi(t)\}$ ; a *linear* autoregression analysis would here give rise to an infinite sequence of residuals, and to a canonical representation which is more complicated than the simple relation (287).

The above argument shows that there is a certain risk of overestimating the outcome of a linear autoregression analysis. When proceeding to residuals of higher order, more parameters are introduced, and it may be that a simpler, possibly non-linear approach would give better results. As mentioned before, the use of non-linear methods does not fall within the scope of this study.

### 31. Some applications of the scheme of moving averages.

In economic theory, a great deal of interest has recently been paid to the schemes of linear regression. However, the discussion of the advantages of the new ideas over the hypothesis of strict periodicity seems to have been carried on exclusively by general theoretical argumentation, without attempting to fit the recommended schemes directly to observational data. Considering, in particular,

the scheme of moving averages, this has not so far as I know been tried on empirical time series in other fields of scientific research either. When selecting statistical material in order to give applications of the previous analysis, I chose economic time series for one reason because of the lack pointed out. In connexion with the account of these applications given in the sequel, we shall touch upon some related lines of economic research where the previous developments seem to yield proper tools for a deeper analysis.

As mentioned in section 9, J. BARTELS (1935) has found »quasi-persistent periodicity» in certain geophysical time series. A periodogram analysis here being inadequate, these series invite an application of the schemes of linear regression. Thanks to their at once flexible and simple construction, these schemes often seem plausible also *a priori*. A few arguments on this line will be touched upon in the sequel when discussing certain geophysical and other phenomena which from a theoretical viewpoint might be interpreted by means of the schemes of linear regression.

The series of yearly wheat prices in Western Europe 1518—1869 compiled by SIR W. BEVERIDGE (1921) was chosen for my earliest experiment in applying the correlogram method. The purpose being to apply a stationary scheme, the analysis was concerned with BEVERIDGE's trend-free index of fluctuations (p. 449 ff.). In order to avoid changes in the structure of the index, the analysis was restricted to the last hundred data. An account of the analysis follows.

Having made the inconsequential modification of reducing the BEVERIDGE index by 100 units, the time series investigated is given in col. (2) of table 7. The first 15 serial coefficients obtained from this material with the use of formula (13) read as follows.

*Table 6. Serial coefficients of the BEVERIDGE wheat price index  
1770—1869.*

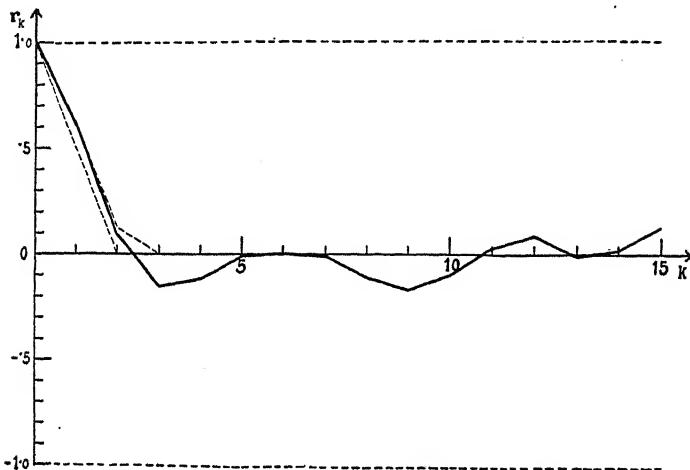
$$\begin{array}{lllll} \bar{r}_1 = .614, & \bar{r}_2 = .090, & \bar{r}_3 = -.156, & \bar{r}_4 = -.115, & \bar{r}_5 = -.006, \\ \bar{r}_6 = .003, & \bar{r}_7 = -.006, & \bar{r}_8 = -.116, & \bar{r}_9 = -.166, & \bar{r}_{10} = -.102, \\ \bar{r}_{11} = .033, & \bar{r}_{12} = .084, & \bar{r}_{13} = -.011, & \bar{r}_{14} = .021, & \bar{r}_{15} = .136. \end{array}$$

The correlogram based on these coefficients is shown in fig. 7.

It is seen that  $\bar{r}_1$  is rather large, and that all of the following serial coefficients are lying in the interval  $-.17 < \bar{r}_k < .14$ , i. e. rather close to zero. To my eye, the correlogram definitely suggests

a scheme of moving averages. Accordingly, the next step in the analysis will be to search for a moving average with autocorrelation coefficients approximating the serial coefficients under investigation. As mentioned before, this problem is due to H. CRAMÉR (see p. 123).

Quite generally, the problem before us may be stated as follows. A set of numbers  $u_1, u_2, \dots, u_h$  being given, does there exist a moving average (249) with autocorrelation coefficients  $r_k$  such that  $r_k = u_k$  for  $1 \leq k \leq h$ ? If the answer is in the affirmative, we know from section 26 that there in general will exist a finite group of moving



*Fig. 7. Correlogram of the BEVERIDGE wheat price index 1770—1869 (thick line), and hypothetical correlograms corresponding to the approaches (292) and (294) (broken lines).*

averages with the prescribed autocorrelation coefficients, and we are also in possession of a direct method for determining the coefficients ( $b$ ) of these moving averages.

Paying regard to the relation (256), we conclude that if there exists a moving average (249) satisfying our conditions, we must have

$$(288) \quad u(x) = u_h x^h + u_{h-1} x^{h-1} + \dots + u_1 x + 1 + \frac{u_1}{x} + \dots + \frac{u_{h-1}}{x^{h-1}} + \frac{u_h}{x^h} =$$

$$= \frac{1}{K^2} (x^h + b_1 x^{h-1} + \dots + b_{h-1} x + b_h) \left( b_h + \frac{b_{h-1}}{x} + \dots + \frac{b_1}{x^{h-1}} + \frac{1}{x^h} \right).$$

If  $x_0$  is a root of the equation  $u(x) = 0$ , then  $1/x_0$  is another

root. It follows that the substitution  $z = x + x^{-1}$  will transform  $u(x)$  into a real polynomial in  $z$  of order  $h$ , say  $v(z)$ . Let us write

$$(289) \quad v(z) = v_0 z^h + v_1 z^{h-1} + \cdots + v_{h-1} z + v_h.$$

The successive calculation of  $v_0, v_1, \dots$  from the coefficients  $u_i$  needs no comment.

It is evident that if  $z$  is a root of  $v(z) = 0$ , two roots of  $u(x) = 0$  will be obtained from the equation

$$(290) \quad P(x, z) = x^2 - zx + 1 = 0.$$

The roots of this equation are given by

$$(291) \quad \frac{z}{2} \pm \sqrt{\frac{z^2}{4} - 1}.$$

The product of the roots being unity, we conclude that unless both the roots are of modulus unity, one of them is situated inside, and the other outside the unit circle.

Denoting conjugate complexity by an asterisk, we know that if  $z$  is a complex root of  $v(z) = 0$ , another root reads  $z^*$ . Further, if  $P(x, z) = 0$  has the roots  $x$  and  $1/x$ , it is evident that  $P(x, z^*) = 0$  has the roots  $x^*$  and  $1/x^*$ . In that case one of the real polynomials  $(x - x_1)(x - x_1^*)$  and  $[x - (x_1)^{-1}] \cdot [x - (x_1^*)^{-1}]$  must be a factor in the polynomial

$$b(x) = x^h + b_1 x^{h-1} + \cdots + b_{h-1} x + b_h$$

appearing in (288).

In case  $v(z) = 0$  presents a real root, say  $z_0$ , we must distinguish two cases. If  $|z_0| \geq 2$ , both  $x_1$  and  $x_2$  are seen to be real. Keeping in mind that either  $x - x_1$  or  $x - x_2$  is a factor in  $b(x)$ , we conclude that this case corresponds to the real roots of the equation  $b(x) = 0$ .

On the other hand, if  $|z_0| < 2$ , we know from (291) that  $x_1$  and  $x_2$  are conjugate complex, and of modulus unity. The factors  $(x - x_1)$  and  $(x - x_2)$  being complex, we conclude that both of them must be contained in  $b(x)$ . Since one zero of  $u(x)$  corresponds to one zero of  $v(x)$ , this is impossible unless  $z_0$  is a root of even multiplicity of  $v(z) = 0$ .

After these remarks, the following theorem demands no explanation.

*Theorem 12.* A necessary and sufficient condition that there exists a moving average (249) with autocorrelation coefficients  $r_k$  equaling  $u_k$  for  $1 \leq k \leq h$  is that the auxiliary polynomial  $v(z)$  defined by (289) has no zero  $z_0$  of odd multiplicity in the real interval  $-2 < z_0 < 2$ .

If this condition is satisfied, the sequences (b) sought for will be given by the coefficients in the real polynomials  $b(x)$  satisfying the relation (288). In full agreement with the analysis in section 26, we conclude further from the above discussion that there are at most  $2^h$  such sequences (b), and that the polynomials  $b(x)$  may be written on the form  $(x - x_1)(x - x_2)\dots(x - x_h)$ , where — denoting by  $z_1, z_2, \dots, z_h$  the roots of  $v(z) = 0$  — the real or complex quantity  $x_1$  is a root of  $P(x, z_1) = 0$ , and  $x_2$  is a root of  $P(x, z_2) = 0$ , etc.

Returning to the BEVERIDGE index of wheat prices, we shall give a few applications of the method outlined above.

In the correlogram  $\bar{r}_k$  (see fig. 7), the small deviations from zero for  $k > 1$  might perhaps be looked upon as pure chance products. Thus we are led to investigate whether there exists a moving average  $\eta(t) + b_1 \eta(t-1)$  with autocorrelation coefficient  $r_1$  equalling .614. Putting  $h = 1$ , and  $u_1 = .614$ , and following the general method, we get  $u(x) = .614x + 1 + .614x^{-1}$ , and  $v(z) = .614z + 1$ . Since the root  $-1/.614 = -1.63$  of  $v(z) = 0$  is lying in the critical interval  $-2 < z < 2$ , we conclude from theorem 12 that there exists no moving average with  $r_1 = .614$  and  $r_k = 0$  for  $k > 1$ .

A short reflection shows that in all moving averages of type  $\eta(t) + b_1 \eta(t-1)$  we have  $-.5 \leq r_1 \leq .5$ . It is further evident that there is only one average of this type such that  $r_1 = .5$ , viz.

$$(292) \quad \zeta(t) - m = \eta(t) + \eta(t-1),$$

Consequently, this average — the correlogram of which is shown in fig. 7 — will yield the closest fit to the prescribed value  $\bar{r}_1 = .614$ . If we rest satisfied with the simple average (292), we have to interpret the deviations between the serial coefficients  $\bar{r}_k$  given on p. 151 and the values  $r_1 = .5, r_2 = r_3 = \dots = 0$  as due to chance. If a better fit is desired, averages involving more parameters (b) must be used. A few examples of this will be given.

As is readily verified, the approach  $u_1 = .614, u_2 = .090$  gives  $u(x) = .090x^2 + .614x + 1 + .614x^{-1} + .090x^{-2}$ , and  $v(z) = .090z^2 + .614z + .820$ . The roots of  $v(z) = 0$  being  $z_1 = -1.82$  and

$z_2 = -5.00$ , it follows from theorem 12 that there does not exist any moving average with the autocorrelation coefficients prescribed. However, only a slight modification in  $u_2$  is sufficient to remove  $z_1$  from the critical interval. In fact, expressing that a root (291) shall equal  $-2$ , we get  $(\frac{1}{2} \bar{r}_1 \cdot u_2^{-1} - 2)^2 = (\frac{1}{2} \bar{r}_1 \cdot u_2^{-1})^2 - (1 - 2 u_2) u_2^{-1}$ , which gives  $u_2 = \bar{r}_1 - \frac{1}{2} = .114$ .

The function  $v(z)$  corresponding to  $u_1 = .614$ ,  $u_2 = .114$  reads  $.114 z^2 + .614 z + .772 = 0$ , and we get  $z_1 = -2.000$ ,  $z_2 = -3.386$ . In order to prepare the construction of the corresponding sequences (b), we solve  $P(x, -2) = x^2 + 2x + 1 = 0$ , which gives the double root  $x = -1$ , and  $P(x, -3.386) = x^2 + 3.486x + 1 = 0$ , which gives the real roots  $x = -3.269$ , and  $x = -3.0591$ . It follows that there exist two binomials  $b(x)$  which satisfy the conditions laid down, viz.  $b_1(x) = (x+1)(x+3.269) = x^2 + 1.3269x + 3.269$ , and  $b_2(x) = (x+1)(x+3.0591) = x^2 + 4.0591x + 3.0591$ .

Using the terminology introduced in section 26, the binomial  $b_1(x)$  gives rise to a regular moving average, viz.

$$(293) \quad \zeta_1(t) - m = \eta(t) + 1.3269 \eta(t-1) + 3.269 \eta(t-2).$$

Since there is only one more process in the group ( $\zeta$ ) of averages with the same correlogram as  $\zeta_1(t)$ , it is evident that this one is symmetrical with  $\zeta_1(t)$ , and thus given by

$$\zeta_2(t) - m = 3.269 \eta(t) + 1.3269 \eta(t-1) + \eta(t-2).$$

In full agreement with the general theory,  $\zeta_2(t)$  can alternatively be derived from  $b_2(x)$  by multiplying the coefficients by  $K_1/K_2$ , taking for  $K_1^2$  the sum of the squared coefficients in  $b_1(x)$ , and similarly for  $K_2^2$  — a short calculation will verify that  $K_1/K_2 = .3269$ , etc.

To check the coefficients (b) obtained, it is sufficient to compute the autocorrelation coefficients of  $\zeta_1(t)$  from the general formula (250). As it should be, we find  $r_1 = .614$ ,  $r_2 = .114$ .

Observing that the increase in  $u_2$  from  $.090$  to  $.114$  has brought on a decrease in  $z_1$  from  $-1.82$  to  $-2.00$ , and an increase in  $z_2$  from  $-5.00$  to  $-3.06$ , it is seen that the parameters are very susceptible to variations in the initial  $u$ -values. Another example of this is given by the fact that a second slight increase in  $u_2$  will make the roots  $z_1$  and  $z_2$  coincide. As follows from (291), this

will occur when  $u_2$  satisfies the relation  $(307)^2 = u_2(1 - 2u_2)$ , i.e. when  $u_2 = .1260$ .

Using the new value  $u_2 = -.1260$ , and taking as before  $u_1 = r_1 = .614$ , we get  $v(z) = .1260z^2 + .614z + .7480$ . The roots of  $v(z) = 0$  read  $z_1 = z_2 = -.614/.252 = -.24365$ , and those of  $P(x_1, -.24365) = 0$  are  $x = -.5225$  and  $x = -.9140$ . Consequently, the binomial  $b(x)$  corresponding to the regular one among the moving averages sought for is simply  $(x + .5225)^2 = x^2 + .10450x + .2730$ . Thus the regular moving average reads

$$(294) \quad \zeta_1(t) - m = \eta(t) + .10450\eta(t-1) + .2730\eta(t-2).$$

Because of the symmetry, it follows that the moving average corresponding to  $b(x) = (x + .9140)^2$  is given by

$$(295) \quad \zeta_2(t) - m = .2730\eta(t) + .10450\eta(t-1) + \eta(t-2).$$

The group  $(\zeta)$  in this case consists of three processes. The remaining one is obtained from  $b(x) = (x + .5225)(x + .9140) = x^2 + .24365x + 1$ . A short calculation gives for this process

$$(296) \quad \zeta_3(t) - m = .5225\eta(t) + .2730\eta(t-1) + .5225\eta(t-2).$$

Checking the calculations, we find that the autocorrelation coefficients of each of the processes (294)–(296) are given by  $r_1 = .6140$ ,  $r_2 = .1260$ . The correlogram of the group (294)–(296) is shown in fig. 7.

The appreciable effect of even small changes in  $u_2$  is evident from the above. Comparing the regular averages (293) and (294), it is seen that the increase in  $u_2$  from .114 to .126 has caused a decrease in  $b_1$  from 1.3269 to 1.0450, and a decrease in  $b_2$  from .3269 to .2730.

The following example shows in detail how the method works when  $v(z) = 0$  has complex roots. Let us start from the values  $u_1 = .60$ ,  $u_2 = .09$ ,  $u_3 = -.15$ ,  $u_4 = -.10$ , which are seen to approximate very closely the first four serial coefficients in the index under investigation. With but little reduction we obtain  $-10^3 v(z) = 10z^4 + 15z^3 - 49z^2 - 105z - 62$ . Solving  $v(z) = 0$ , we get  $z_1 = -.21272$ ,  $z_2 = .25103$ ,  $z_3 = -.9415 + .5240i$ ,  $z_4 = -.9415 - .5240i$ . Concluding from theorem 12 that there exists a group of moving averages with the prescribed correlogram, a short reflection shows that the group consists of 8 processes.

Solving  $P(x, z_1) = 0$ , we get  $x = -\cdot7013$ , and  $x = -1\cdot4259$ , while  $P(x, z_2)$  gives  $x = \cdot4966$  and  $x = 2\cdot0137$ . The roots of  $P(x, z_3) = 0$  were found to be  $x = -\cdot3381 - \cdot6679i$ , and  $x = -\cdot6034 + + 1\cdot1919i$ . According to the discussion preceding theorem 12, the roots of  $P(x, z_4) = 0$  are obtained by replacing  $i$  by  $-i$  in the roots of  $P(x, z_3) = 0$ .

Writing

$$B(x) = (x + \cdot3381 - \cdot6679i)(x + \cdot3381 + \cdot6679i) = x^2 + \\ + \cdot6762x + \cdot560402,$$

the regular moving average will be obtained from  $b(x) = (x + \cdot7013)(x - \cdot4966) \cdot B(x) = x^4 + \cdot8809x^3 + \cdot3505x^2 - \cdot1208x - - \cdot1952$ , and reads

$$(297) \quad \eta(t) + \cdot8809\eta(t-1) + \cdot3505\eta(t-2) - \cdot1208\eta(t-3) - \\ - \cdot1952\eta(t-4).$$

Squaring the coefficients, we get the sum  $K^2 = 1\cdot95153$ .

According to the general analysis, a second average with the same correlogram will be delivered by  $b_1(x) = (x + 1\cdot4259)(x - \cdot4966) \cdot B(x) = x^4 + 1\cdot6055x^3 + \cdot4807x^2 + \cdot0420x - \cdot3968$ . The sum of squared coefficients is  $K_1^2 = 3\cdot967917$ , which gives  $K/K_1 = \cdot701304$ . Multiplying the coefficients in  $b_1(x)$  by this factor, we get the coefficients in the corresponding moving average. This is found to be

$$\cdot7013\eta(t) + 1\cdot1259\eta(t-1) + \cdot3371\eta(t-2) + \cdot0294\eta(t-3) - \\ - \cdot2783\eta(t-4).$$

A third moving average in the group is seen to be obtained from  $(x + \cdot7013)(x - 2\cdot0137) \cdot B(x)$ . Proceeding as before, we find for the corresponding process

$$\cdot4966\eta(t) - \cdot3159\eta(t-1) - \cdot8637\eta(t-2) - \cdot8395\eta(t-3) - \\ - \cdot3930\eta(t-4).$$

In the same way, the polynomial  $(x + 1\cdot4259)(x - 2\cdot0137) \cdot B(x)$  yields a fourth process in the group, viz.

$$\cdot3483\eta(t) + \cdot0308\eta(t-1) - \cdot9432\eta(t-2) - \cdot7909\eta(t-3) - \\ - \cdot5604\eta(t-4).$$

The four remaining averages in the group considered correspond to the complex roots  $x = -\cdot6034 \pm 1\cdot1919 i$  of  $b(x) = 0$ . Due to the symmetry, these processes may be obtained directly from the four processes above by reversing the order of the coefficients. For instance, the regular average (297) gives

$$-\cdot1952 \eta(t) - \cdot1208 \eta(t-1) + \cdot3505 \eta(t-2) + \cdot8809 \eta(t-3) + \eta(t-4).$$

The above computations have been checked by verifying that the autocorrelation coefficients of the four processes equal the prescribed values  $r_1 = \cdot6000$ ,  $r_2 = \cdot0900$ ,  $r_3 = -\cdot1500$ ,  $r_4 = -\cdot1000$ .

Having now exemplified the construction of sets (b) belonging to moving averages (249) with correlograms approximating that of the BEVERIDGE wheat price index, we shall postpone the discussion of the results arrived at until we have made a few applications of the inversion formula (200) and the relation (260).

If  $\{\zeta(t)\}$  is a regular moving average (249), the primary process  $\{\eta(t)\}$  will be given either by (200) or by (255), the latter formula corresponding to the exceptional case when the characteristic equation of the difference relations satisfied by the coefficients (a) presents roots of modulus unity. This characteristic equation is nothing else than the equation  $b(x) = 0$  used in the general method exemplified above. This method being based on the calculation of the roots of the equation mentioned, we are in a position to point out directly which of the formulae (200) and (255) to apply in the different examples, and to carry the analysis further on the basis of the general developments in section 26.

Returning first to the approach (292), we have to apply (255), for the root  $x = -1$  of  $b(x) = 0$  is of modulus unity. Proceeding as in the illustration 1 of section 26, we get

$$(298) \quad \eta(t) = \lim_{1 > a \rightarrow -1} [\zeta(t) - m - a(\zeta(t-1) - m) + a^2(\zeta(t-2) - m) - \dots]$$

By construction, the second approach (293) is also such that a root of  $b(x) = 0$  is of modulus unity.

In order to get an application of (200), we proceed to the approach (294). According to theorem 8, the coefficients (a) will satisfy the difference relation  $a_k + 1\cdot0450 a_{k-1} + \cdot2730 a_{k-2} = 0$ . The characteristic equation being  $b(x) = (x + \cdot5225)^2 = 0$ , it follows from section 6 that  $a_k$  may be written on the form  $a_k = (A + k \cdot B) \cdot (-\cdot5225)^k$ . The constants  $A$  and  $B$  may be obtained from the

initial values  $a_0 = 1$ ,  $a_1 = -1.0450$  (cf. (97) and (197)). A short calculation gives  $A = B = 1$ , and hence  $a_k = (1 + k)(-5225)^k$ . In applying formula (200), we have to insert this expression, and to replace  $\{\zeta(t)\}$  by  $\{\zeta_1(t) - m\}$ , where  $m = E[\zeta_1(t)]$ . Observing that  $\sum_0^{\infty} a_k = (1.5225)^{-2} = .4314$ , we get the inversion formula

$$(299) \quad \eta(t) = -.4314 m + \zeta_1(t) - 2(5225) \zeta_1(t-1) + \\ + 3(5225)^2 \zeta_1(t-2) - 4(5225)^3 \zeta_1(t-3) + \dots$$

It is seen that the approach (297) gives a formula of the same type, but, of course, with more complicated coefficients ( $a$ ).

Denoting the BEVERIDGE index by  $\xi_t$ , and assuming that  $\xi_t$  is a sample series of the moving average  $\{\zeta_1(t)\}$  given by (294), formula (299) may be used for deriving a series  $\bar{\eta}_t$  which corresponds to this hypothesis. Identifying  $m$  with the average  $\bar{m}$  of the index series  $\bar{\eta}_t$ , we get

$$(300) \quad \bar{\eta}_t = -.4314 \bar{m} + \xi_t - 1.0450 \xi_{t-1} + .8190 \xi_{t-2} - .5706 \xi_{t-3} + \dots$$

The sum of the 100 index fluctuations  $\xi_t$  given in table 7 being  $-28$ , we get  $\bar{m} = -.28$ . Inserting this, and using the index deviations given in col. (1), formula (300) has given the series  $\bar{\eta}_t$  presented in col. (2). Having put  $a_k = 0$  for  $k > 12$ , the first 12  $\bar{\eta}_t$ -values are partly based on index fluctuations not given in the table.

Apart from the constant  $\bar{m} = -.28$ , the moving average (294) as performed on col. (2) must reproduce col. (1). Thus the values  $\bar{\eta}_t$  obtained may be checked by the simple identity

$$(301) \quad \xi_t - \bar{m} = \bar{\eta}_t + 1.0450 \bar{\eta}_{t-1} + .2730 \bar{\eta}_{t-2}.$$

Because of the large number of terms in (300) and (301), there will sometimes be a deviation amounting to  $.2$  or  $.3$ .

In analogy with the above, the series  $\bar{\eta}_t$  corresponding to the regular approach (292) may be obtained by a limit passage based on the relation (255). Having exemplified in illustration 1 of section 26 the limit procedure for deriving a primary series  $\bar{\eta}_t$ , it is seen that in point of principle no complications will be met. Accordingly, we shall not dwell further on the exceptional cases when  $b(x) = 0$  presents a root of modulus unity.

Starting from the hypothesis that a given time series  $\xi_t$  is a sample series of a regular moving average (249), we have above

Table 7. BEVERIDGE wheat price index fluctuations (col. 1), and hypothetical primary series  $\bar{\eta}_t$  (col. 2).

Year	(1)	(2)									
1770	31	23.9	1795	30	17.6	1820	-16	.8	1845	15	19.6
71	36	9.3	96	-5	-27.2	21	-24	-19.3	46	39	19.9
72	19	3.0	97	-16	7.8	22	-23	-2.8	47	-10	-35.9
73	6	.5	98	-13	-13.5	23	-29	-20.5	48	-20	12.3
74	5	3.9	99	20	32.2	24	-29	-6.4	49	-26	-28.8
75	-12	-15.9	1800	39	9.3	25	-31	-18.3	1850	-22	5.0
76	-16	-2	01	17	-1.2	26	-18	3.2	51	-14	-11.1
77	-6	-1.2	02	5	4.0	27	-7	-5.1	52	5	15.5
78	-13	-11.4	03	-6	-9.6	28	14	18.7	53	38	25.1
79	-21	-8.5	04	25	34.2	29	3	-14.5	54	41	10.8
1780	-13	-7	05	14	-18.9	1830	10	20.7	55	38	20.1
81	-12	-8.6	06	-2	8.6	31	5	-12.3	56	7	-16.7
82	-6	3.5	07	-7	-10.6	32	-18	-10.6	57	-18	-5.8
83	-6	-6.9	08	-6	3.1	33	-20	-5.4	58	-19	-8.1
84	-8	-1.3	09	-6	-6.0	34	-22	-13.3	59	-3	7.4
85	-15	-11.4	1810	4	9.8	35	-18	-2.5	1860	16	10.7
86	-16	-3.4	11	40	31.8	36	-12	-5.6	61	7	-6.0
87	-7	.0	12	21	-14.5	37	2	8.7	62	-8	-4.4
88	8	9.2	13	-4	2.9	38	17	9.6	63	-21	-14.5
89	8	-1.4	14	-4	-2.7	39	7	-5.2	64	-19	-2.4
1790	-14	-14.8	15	30	32.3	1840	-5	-1.9	65	-6	.7
91	-22	-5.9	16	78	45.2	41	1	4.8	66	19	19.3
92	-13	-2.6	17	26	-29.7	42	-8	-12.1	67	18	-1.9
93	-15	-10.5	18	-6	18.1	43	-12	-3	68	-7	-9.8
94	3	14.9	19	-14	-19.3	44	-8	-4.1	69	2	13.1

exemplified a general method for deriving the corresponding primary series  $\bar{\eta}_t$ . We are now in a position to solve the analogous problem under the hypothesis that  $\xi_t$  is a sample series of a non-regular average (cf. p. 129). After the detailed treatment of the regular case, it will be sufficient to deal with the non-regular averages very briefly.

Let the characteristic equation of a non-regular moving average  $\{\zeta(t)\}$  be given by (cf. (258))

$$b(x) = \frac{K}{K^{(i)}}(x^h + b_1^{(i)}x^{h-1} + \cdots + b_{h-1}^{(i)}x + b_h^{(i)}) = \\ = \frac{K}{K^{(i)}}(x^p + A_1 x^{p-1} + \cdots + A_{p-1}x + A_p)(x^q + B_1 x^{q-1} + \cdots + \\ + B_{q-1}x + B_q) = 0,$$

where  $p + q = h$ , and where the roots of  $x^p + A_1 x^{p-1} + \cdots + A_p = 0$  are lying inside, and those of  $x^q + B_1 x^{q-1} + \cdots + B_q = 0$  outside the unit circle.

The coefficients  $A_i$  and  $B_i$  being real, two real sequences  $p_i$  and  $q_i$  will be defined by the systems (cf. (97))

$$\begin{cases} p_n + A_1 p_{n-1} + \cdots + A_{n-1} p_1 + A_n = 0, & n = 1, 2, \dots, p-1; \\ p_n + A_1 p_{n-1} + \cdots + A_{p-1} p_{n-p+1} + A_p p_{n-p} = 0, & n \geq p; \end{cases}$$

$$\begin{cases} B_q q_n + B_{q-1} q_{n-1} + \cdots + B_{q-n+1} q_1 + B_{q-n} = 0, & n = 1, 2, \dots, q-1; \\ B_q q_n + B_{q-1} q_{n-1} + \cdots + B_1 q_{n-q+1} + q_{n-q} = 0, & n \geq q; \end{cases}$$

$p_0 = q_0 = 1$ . It is seen that the sum

$$\alpha(t) = \zeta(t) + p_1 \zeta(t-1) + p_2 \zeta(t-2) + \cdots$$

will be convergent. Paying regard to certain evident relations between the coefficients  $A_k$ ,  $B_k$ , and  $b_k^{(i)}$ , it follows further

$$(302) \quad \alpha(t) = K \cdot [\eta(t) + B_1 \eta(t-1) + \cdots + B_q \eta(t-q)] / K^{(i)}.$$

Thus prepared, let us form

$$\beta(t) = \alpha(t) + q_1 \alpha(t+1) + q_2 \alpha(t+2) + \cdots$$

Observing that the roots of the equation

$$B_q x^q + B_{q-1} x^{q-1} + \cdots + B_1 x + 1 = 0$$

are of modulus less than unity, it follows without difficulty that this sum also is convergent, and that  $\beta(t) = B_q \cdot K \cdot \eta(t-q) / K^{(i)}$ . Since  $B_h \neq 0$ , we conclude that, apart from a constant factor, the repeated linear operation performed will yield  $\{\eta(t)\}$  in terms of the non-regular moving average considered.

Comprehending the double transformation, we get

$$(303) \quad \{\eta(t)\} = \frac{K^{(i)}}{B_q \cdot K} \cdot \sum_{n=-\infty}^{\infty} c_n \cdot \{\zeta(t+q+n)\},$$

where  $c_0 = 1 + p_1 q_1 + p_2 q_2 + \dots$ , and

$$\begin{cases} c_n = p_n + p_{n+1} q_1 + p_{n+2} q_2 + \dots, & n > 0, \\ c_{-n} = q_n + q_{n+1} p_1 + q_{n+2} p_2 + \dots, & n > 0. \end{cases}$$

The generalization to the case when  $b(x) = 0$  presents roots on the periphery of the unit circle is, of course, straightforward.

In order to give an explicit application of the relation (303), let us consider the group  $(\zeta)$  belonging to the regular average (294) studied in detail before. As to (295), we get by analogy from (299), and in full agreement with (303),

$$(304) \quad \eta(t-2) + .4314 m = \zeta_2(t) - 2(-.5225)^2 \zeta_2(t+1) + \\ + 3(-.5225)^3 \zeta_2(t+2) - \dots$$

Considering the remaining average (296) in the group, the coefficients appearing in (303) are seen to reduce to  $p_n = q_n = (-.5225)^n$ . Hence  $c_n = p^{|n|}/(1-p^2)$ ,  $p = -.5225$ ,  $n \geq 0$ . Since  $K/K^{(t)}$  is nothing else than the factor of  $\eta(t)$  in (296), and  $B_q = -1/.5225$ , we get

$$(305) \quad \eta(t-1) = \sum_{n=-\infty}^{\infty} [\zeta_3(t+n) - m] \cdot p^{|n|}/(1-p^2), \quad p = -.5225.$$

By means of the formulae (304) and (305), it is possible to derive explicitly the two series  $\bar{\eta}_t$  corresponding to the hypotheses that the BEVERIDGE index is a moving average (295) or (296) respectively. The calculations running as in the regular case, no detailed illustration will be given. Of course, the last  $\bar{\eta}_t$ -values can be calculated only approximately. It should be observed that a complete check may be based on identities similar to (301).

If a series  $\bar{\eta}_t$  corresponding to a certain process in a group  $(\zeta)$  has been derived, it will sometimes be possible to arrive at the primary series, say  $\bar{\eta}_t^*$ , corresponding to another average in the group by means of a simpler procedure than that based on (303). For instance, representing by  $\bar{\eta}_t$  and  $\bar{\eta}_t^*$  the primary series corresponding to (294) and (296) respectively, the relation (260) gives

$$-\bar{\eta}_t = p \bar{\eta}_t^* - (1-p^2) \bar{\eta}_{t-1}^* - p(1-p^2) \bar{\eta}_{t-2}^* - p^2(1-p^2) \bar{\eta}_{t-3}^* - \dots$$

where  $p = -.5225$ , and hence  $p \bar{\eta}_t^* - \bar{\eta}_{t-1}^* = p \bar{\eta}_{t-1} - \bar{\eta}_t$ . The series  $p \bar{\eta}_{t-1} - \bar{\eta}_t$ , say  $\bar{\xi}_t^*$ , being readily derived from the series  $\bar{\eta}_t$  in table 7, we get the simple relation (cf. (305))

$$\bar{\eta}_{t-1}^* = -\bar{\zeta}_t^* + p \bar{\zeta}_{t+1}^* - p^2 \bar{\zeta}_{t+2}^* + p^3 \bar{\zeta}_{t+3}^* - \dots$$

Having now given an account of some applications of the general theory of moving averages to the BEVERIDGE wheat price index, we shall next attach some remarks of general scope to certain points in the analysis. Let us in the first place touch upon the problem of testing the results derived under a hypothesis of moving averages.

In a process  $\{\zeta(t)\}$  of moving averages as given by (249), a characteristic feature is that  $\zeta(t)$  is independent of  $\zeta(t-h-k)$  for  $k > 0$ .

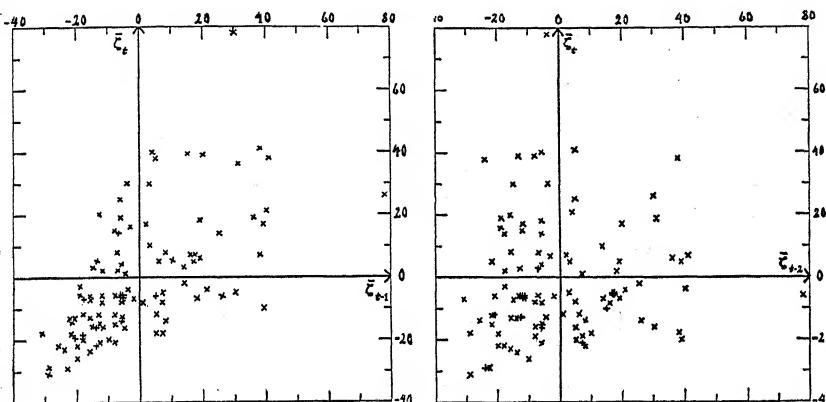


Fig. 8. BEVERIDGE wheat price index,  $\bar{\zeta}_t$ . The scatters  $(\bar{\zeta}_t, \bar{\zeta}_{t-1})$  (left), and  $(\bar{\zeta}_t, \bar{\zeta}_{t-2})$  (right).

Thus, (249) will form no adequate basis for the analysis of a time series  $\zeta_t$  unless the scatters  $(\bar{\zeta}_t, \bar{\zeta}_{t-h-1})$ ,  $(\bar{\zeta}_t, \bar{\zeta}_{t-h-2})$ , etc. approximate distributions of independent variables. On this line it would, of course, be possible to develop different types of tests. — Figure 8 contains the scatters  $(\bar{\zeta}_t, \bar{\zeta}_{t-1})$  and  $(\bar{\zeta}_t, \bar{\zeta}_{t-2})$  belonging to the BEVERIDGE data given in table 7. Figure 8 (left) clearly shows that  $\bar{\zeta}_t$  and  $\bar{\zeta}_{t-1}$  must be considered interdependent. On the other hand, the scatter  $(\bar{\zeta}_t, \bar{\zeta}_{t-2})$  seems to permit us to look upon  $\bar{\zeta}_t$  and  $\bar{\zeta}_{t-2}$  as independent, a circumstance speaking in favour of an average of the simple type  $\bar{\zeta}_t = \bar{\eta}_t + b \bar{\eta}_{t-1}$ .

In starting from a specified scheme of moving averages, we are in a position to derive a hypothetical value for any characteristic of the series  $\bar{\zeta}_t$  and of the corresponding series  $\bar{\eta}_t$  calculated as indicated above. In point of principle, every such characteristic as

compared with the corresponding empirical one will give a basis for testing the hypothetical set-up.

Of course, when a scheme with correlogram approximating the empirical one is chosen, there will automatically be an agreement in the main between certain hypothetical and empirical characteristics. For instance, the serial coefficients of  $\bar{\eta}_t$  will approximate zero, since the deviations from this hypothetical value will be due only to the differences between the hypothetical and the empirical correlogram. The situation is the same with regard to the ratio between the variances of  $\bar{\eta}_t$  and  $\bar{\zeta}_t$ . Considering e.g. the approach (294), the hypothetical value of this quotient is  $1/(1 + 1 \cdot 0450^2 + 2730^2) = 462$ . On the other hand, the variances of the series  $\bar{\eta}_t$  and  $\bar{\zeta}_t$  being 213.7 and 383.6 respectively, the empirical quotient equals .557.

By construction, the moving averages in a group  $(\zeta)$  will present the same correlogram and the same variance. We conclude from the above that in point of principle the autocorrelation properties of the corresponding series  $\bar{\eta}_t$  will give us no criterion for distinguishing which of the processes should be preferred. Expressing the argument in other words, the deviations in  $\bar{r}_k(\bar{\eta}_t)$  and  $\bar{D}(\bar{\eta}_t)$  from the hypothetical values will depend solely on the differences between the hypothetical and empirical correlograms of  $\bar{\zeta}_t$ , and these differences are exactly equal for all processes in the group  $(\zeta)$ .

The forecasts based on the hypothesis of moving averages disclose an interesting aspect of the test problem. Denoting by  $F_t[\zeta(t+k)]$  the forecast over  $k$  time units based on the development up to the time point  $t$ , the general formula (213) gives in the approach (294) the following forecasts

$$(306) \quad F_t[\zeta(t+1)] = \bar{m} + 1 \cdot 0450 \bar{\eta}_t + 2730 \bar{\eta}_{t-1};$$

$$F_t[\zeta(t+2)] = \bar{m} + 2730 \bar{\eta}_t;$$

and  $F_t[\zeta(t+k)] = \bar{m}$  for  $k > 2$ . In particular, taking  $t = 1811$ , table 7 shows that  $\bar{\eta}_t = 31.8$ ,  $\bar{\eta}_{t-1} = 9.8$ . Keeping in mind that  $\bar{m} = -28$ , we get  $F_t[\zeta(t+1)] = 35.6$ ,  $F_t[\zeta(t+2)] = 2.4$ , while the actual path of the index runs through  $\zeta_{t+1} = 21$ ,  $\zeta_{t+2} = -4$ .

Considering on the other hand the approach (295), which belongs to the same group as (294), the forecast formulae corresponding to (306) would read  $F_t[\zeta(t+1)] = \bar{m} + 1.0450 \bar{\eta}_t + \bar{\eta}_{t-1}$ ;  $F_t[\zeta(t+2)] = \bar{m} + \bar{\eta}_t$ . However, in this case we cannot derive  $\bar{\eta}_t$  and  $\bar{\eta}_{t-1}$

from the observed values  $\xi_t, \xi_{t-1}, \xi_{t-2}, \dots$ . Now, the autoregression analysis as developed in Chapter II yields a linear forecast which is valid for any stationary process with finite dispersion, and for which the squared deviation from the future path is of minimum expectation. Writing  $\{\eta^{(i)}(t)\}$  for the residual process of the stationary process considered, and paying regard to the results arrived at in section 26, this forecast will in the present case reduce to

$$(307) \quad F_t[\zeta(t+k)] = b_k \eta_t^{(i)} + b_{k+1} \eta_{t-1}^{(i)} + \dots + b_h \eta_{t-h+k}^{(i)}$$

where the sequence  $(b)$  is identical with the coefficients in the corresponding regular average. According to the general theory,  $\{\eta^{(i)}(t)\}$  is obtained simply by subjecting  $\{\zeta(t)\}$  to the same linear operation which gives the primary process  $\{\eta(t)\}$  in terms of the regular average.

In taking the squared deviation as the measure of the efficiency of a forecast, we conclude from the above that the different hypothetical averages in a group  $(\zeta)$  will give rise to exactly the same forecast series  $F_t[\zeta(t+1)], F_t[\zeta(t+2)],$  etc. A simple illustration of this is given by the group (294)–(296). Taking e.g. the process (295) for a hypothetical basis, we have in the first place to form the series  $\bar{\eta}_t^{(i)}$ . According to the general analysis, this is identical to col. (2) in table 7. Applying next the general relation (307), and keeping in mind that the coefficients  $(b)$  coincide with those in the regular average (294), it follows that the resulting forecasts will equal those previously obtained on the basis of the regular average (294).

Expressing the situation in other words, we have found that the different averages in a group  $(\zeta)$  are equivalent in view of those aspects of the general test problem hitherto considered. Since the indeterminateness is due to our having dealt with only the auto-correlation properties of the hypothetical models, we have to use other types of methods for distinguishing between the different averages in a group. Generally speaking, these tests should examine to what degree the elements  $\bar{\eta}_t$  and  $\bar{\eta}_{t+k}$  resulting from a special hypothesis might be considered not only uncorrelated but also independent. The different averages in a group giving rise to different primary series  $\bar{\eta}_t$ , we must therefore compare in detail the multi-dimensional scatters  $(\bar{\eta}_t, \bar{\eta}_{t-1}, \dots, \bar{\eta}_{t-h})$ . For the sake of concreteness, the scatter  $(\bar{\eta}_t, \bar{\eta}_{t-1})$  obtained from col. (2) in table 7 is

shown in fig. 9. To my eye, the scatter forms no very good approximation to a distribution of independent variables. In performing different tests of this nature, it may of course occur that no series  $\bar{\eta}_t$  will give satisfactory results — perhaps one of the scatters examined will suggest instead a non-linear scheme, e.g. a moving average performed on the non-autocorrelated process  $\{\xi(t)\}$  defined by (286). However, it would be beyond the frame of this study to enter upon further details concerning the non-linear schemes and the non-linear methods required for distinguishing between the different moving averages belonging to the same group.

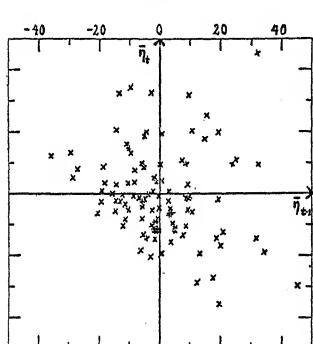


Fig. 9. BEVERIDGE wheat price index. The scatter  $(\bar{\eta}_t, \bar{\eta}_{t-1})$  belonging to a hypothetical primary series  $\bar{\eta}_t$

As repeatedly emphasized in earlier sections, the applications accounted for in the present study do not aim at reaching definitive, quantitative results. Accordingly, no attempts will be made to test the significance of the parameters arrived at under the different hypotheses dealt with. Hence it is out of the question to draw quantitative comparisons with earlier investigations of the wheat price data analysed; as is well known, SIR W. BEVERIDGE (1922) has subjected his index to an extensive periodogram analysis, while G. U. YULE (1926) has illustrated certain new correlogram methods by the use of the BEVERIDGE index numbers. However, it will be illuminating to take up some points of these investigations for a qualitative comparison with the moving average approach.

In his presentation of the wheat price data, SIR W. BEVERIDGE (1921) gives two series of index numbers, the trend present in the first series being removed in the second. In contradistinction to BEVERIDGE and to the present writer, G. U. YULE (1926) works on the first series. In his search for hidden periodicities, YULE modifies the correlogram method because of the trend present in the original data. Since YULE's approach takes into consideration the differences of the series analysed, it might be interpreted as an application of certain non-stationary or evolutive processes of the homogeneous type  $\{\xi(t)\}$  defined by (192). The investigation thus following a quite different line of research, no further comment is called for in the present connexion.

The periodogram method being based on the assumption of hidden periodicities, its use will result in a hypothetical scheme additively built up by a functional element  $y(t)$  consisting of a number of superposed harmonics, and a random element or »error»  $\eta(t)$ . Since a single harmonic involves three parameters, the total number of parameters in a scheme of hidden periodicities amounts to thrice the number of harmonics superposed. Specifying the values for these parameters, we can compute the functional element  $y(t)$ . Deducting  $y(t)$  from the original data, say  $\xi_t$ , we obtain the corresponding path of the random component,  $\bar{\eta}_t = \xi_t - y(t)$ .

A common standard measure of the efficiency of the periodogram analysis is obtained by dividing the variance of the errors  $\bar{\eta}_t$  by the variance of the original data  $\xi_t$ . This quotient, say  $\bar{x}^2$ , is always below unity, and the closer to zero the quotient, the more the functional element  $y(t)$  will »explain» of the series under analysis. For instance, the harmonic corresponding to the largest ordinate in the periodogram 1545—1844 given by SIR W. BEVERIDGE (1922), p. 438) will leave an »error»  $\bar{\eta}_t$  with variance amounting to 91 % of the variance of the wheat price index. The harmonic in question is of period  $p \sim 15.225$ .

On the other hand, a scheme of moving averages (249) is built up by means of a random variable  $\eta(t)$ , here called primary, and a set of coefficients or parameters  $(b)$ . After having chosen numerical values for the parameters  $(b)$ , the general analysis in the present section provides a method for deriving from the original series  $\xi_t$  the corresponding path  $\bar{\eta}_t$  of the primary variable.

Even in the case of moving averages, the quotient  $\bar{x}^2$  between the variances of  $\bar{\eta}_t$  and  $\xi_t$  may be taken as a standard measure of the efficiency of the analysis. The hypothetical value for  $\bar{x}^2$  being in this case  $x^2 = 1/(1 + b_1^2 + \dots + b_n^2)$ , it is seen that the larger the coefficients  $(b)$ , the less is  $x^2$ , and the better the result of the analysis. For instance, considering the  $\bar{\eta}_t$ -values in table 7 derived from the approach (294), we have, as already mentioned,  $\bar{x}^2 = .557$ ,  $x^2 = .462$ . The small values obtained seem rather satisfactory, but it must be remembered that the analysis has been restricted to the last 100 BEVERIDGE data 1770—1869. Thus, although the approach (294) involves only two parameters, it might perhaps be unfair to the scheme of hidden periodicities to compare directly with the  $\bar{x}^2$ -value .91 derived from the whole series under the hypothesis of one harmonic component.

An examination of the forecasts delivered will show clearly the thorough-going difference between the scheme of hidden periodicities and the scheme of moving averages.

In a scheme of hidden periodicities of type (39), the forecast curve is identical to the functional element  $y(t)$ , i.e. the sum of harmonics superposed (cf. p. 59). The hypothetical model thus will provide a definite functional forecast over the infinite future. The expected squared deviation between the actual development and the forecast is independent of the period forecasted over, and amounts to  $D^2(\eta)$ .

A scheme of moving averages gives a quite different type of forecast. In fact, considering a moving average (249) ranging over  $h + 1$  time units, it is only the forecasts over the next  $h$  observations that are effective — the forecasts beyond  $h$  time units are trivial, and reduce to the average of the original data. According to the general analysis, the different averages in a group ( $\zeta$ ) will give rise to the same forecasts, viz.

$$F_t[\zeta(t+k)] = \bar{m} + b_k \bar{\eta}_t + b_{k+1} \bar{\eta}_{t-1} + \cdots + b_h \bar{\eta}_{t-h+k},$$

where the coefficients ( $b$ ) are those appearing in the regular average used as the basis for deriving the primary series  $\bar{\eta}_t$ . The hypothetical value for the squared deviation from the actual development being (cf. (218))

$$(1 + b_1^2 + b_2^2 + \cdots + b_{k-1}^2) \cdot D^2(\eta),$$

it is seen that the efficiency of the prognosis will decrease gradually as the period forecasted over is extended.

Especially in view of economic time series, the type of forecast delivered by the scheme of moving averages seems *a priori* more realistic, seems to correspond better to what might be reasonably possible to find out from the past development. Further, considering the forecasts over a short period, the prognosis given by the scheme of moving averages is, as a rule, rather efficient. In my opinion, this is a circumstance of central importance, for often the main interest is concentrated upon the prognosis concerning the near future.

As to the harmonic components suggested by a periodogram analysis, these cannot always be interpreted in the light of what is otherwise known of the phenomena under investigation. Lacking

other evidence, the periodicities thus will stand out as quite isolated results of the analysis. As pointed out by SIR W. BEVERIDGE ((1922), p. 438), this is the case with the period of length 15·225 years suggested in his periodogram mentioned above. — Against this background, the scheme of moving averages seems more fertile. Let us dwell a moment on this point.

In modern economic-statistical research work, a prominent line of approach is the regression analysis of time series (cf. e.g. C. F. Roos (1934)). Denoting a set of time series by  $\bar{\zeta}_t, \bar{\eta}'_t, \bar{\eta}''_t, \dots, \bar{\eta}_t^{(n)}$ , a simple type of approach reads

$$(308) \quad \bar{\zeta}_t = c_1 \bar{\eta}'_t + c_2 \bar{\eta}''_t + \dots + c_n \bar{\eta}_t^{(n)} + \bar{\varepsilon}_t,$$

where  $\bar{\varepsilon}_t$  stands for the residual in the representation of  $\bar{\zeta}_t$  as linearly correlated with the  $n$  variables  $\bar{\eta}_t^{(i)}$ . A generalized approach is obtained by replacing  $\bar{\eta}_t^{(i)}$  by  $\bar{\eta}_{t-k_i}^{(i)}$ , where the constant  $k_i$  represents the lag of the series  $\bar{\eta}_t^{(i)}$  behind the series  $\bar{\zeta}_t$ . The concept of *distributed lag* implies a further generalization, which in the simplest case  $n = 1$  leads to an approach of type

$$(309) \quad \bar{\zeta}_t = b_0 \bar{\eta}_t + b_1 \bar{\eta}_{t-1} + \dots + b_h \bar{\eta}_{t-h} + \bar{\varepsilon}_t.$$

For instance, as an initial approximation we might represent a wheat price (i.e.  $\bar{\zeta}$ ) in the year  $t$  as linearly correlated with wheat crops (i.e.  $\bar{\eta}$ ) in the years  $t, t-1, \dots, t-h$ . According to the theory of supply and demand, we might expect that in this case the dominant regression coefficients  $b_i$  would be negative.

Disregarding the residual  $\bar{\varepsilon}_t$ , and interpreting in the terminology of the present study, the series  $\bar{\zeta}_t$  as given by (309) is seen to be nothing else than a moving average performed on the primary series  $\bar{\eta}_t$ . Consequently, a hypothetical model corresponding to the approach (309) will be obtained by adding two independent processes, viz. a moving average  $\{b_0 \eta(t) + b_1 \eta(t-1) + \dots + b_h \eta(t-h)\}$  and a residual process  $\{\varepsilon(t)\}$ .

Having seen that the concept of moving average may be attached directly to the concept of *distributed lag*, the theory of moving averages as developed in the present study seems to disclose new aspects of lag problems, and to suggest methods for a deeper analysis in this field of research. A few remarks on this line will follow.

The idea underlying the general methods applied in the present section is that the autocorrelation properties of a time series  $\xi_t$  will reveal whether the process of moving averages is an adequate type of hypothetical model for  $\xi_t$ . The circumstances of importance in this connexion are (A) that in such an analysis no other time series are taken into consideration, and (B) that a specified hypothesis of moving averages (it should be observed that the different averages in a group  $(\zeta)$  must be examined separately!) will determine a hypothetical primary series  $\bar{\eta}_t$ . Accordingly, independently of the first stage of the analysis we can compare such a hypothetical series  $\bar{\eta}_t$  with other time series, say  $\bar{\eta}'_t$ ,  $\bar{\eta}''_t$ , etc., thought of as possibly affecting the series  $\xi_t$  examined. For instance, in applying the approach (294) to the BEVERIDGE wheat price index, we have derived the hypothetical series  $\bar{\eta}_t$  given in col. (2) of table 7. This series has been calculated merely in order to illustrate in detail how the general inversion formula (200) works when applied to an observational time series, but in case the parameters  $(b)$  were significant, the series  $\bar{\eta}_t$  might be compared with e.g. some appropriate wheat crop series  $\bar{\eta}'_t$ . Following the suggestion made in connexion with the approach (309), we might in such a case change simultaneously the signs in the coefficients  $(b)$  and the series  $\bar{\eta}_t$ .

In view of the lines of research suggested above, our theoretical analysis of moving averages calls for generalizations in various directions. Having assumed the primary process  $\{\eta(t)\}$  in (249) to be purely random, it would in the first place be of interest to generalize the concept of moving average by removing the restrictions imposed on the primary process  $\{\eta(t)\}$ . Now, assuming only that  $\{\eta(t)\}$  is stationary and of finite dispersion, the autocorrelation coefficients of  $\{\zeta(t)\}$  will exist, and be obtained by a straightforward generalization of the relations (250). Considering two time series  $\xi_t$  and  $\bar{\eta}'_t$  with correlograms  $r_k(\xi)$  and  $r_k(\bar{\eta}'_t)$  respectively, the generalized relations evidently may be used for finding out approximately how a moving average with specified coefficients  $(b)$  as performed on  $\bar{\eta}'_t$  would transform  $r_k(\bar{\eta}'_t)$ . If the transformed correlogram approximates  $r_k(\xi)$ , we are led to investigate in detail whether  $\xi_t$  approximates a moving average with these coefficients  $(b)$ , and with  $\bar{\eta}'_t$  for primary series. Concluding from theorem 9 that the representations (200) and (303) hold even for the generalized average, this investigation can be performed as suggested in the case of a purely random  $\{\eta(t)\}$ , viz. by deriving directly from  $\xi_t$  the primary series

$\bar{\eta}_t$  corresponding to the coefficients  $(b)$ , and then correlating or comparing in another way the two series  $\bar{\eta}_t$  and  $\bar{\eta}'_t$ .

Finally, another generalization is introduced when several series  $\bar{\eta}_t$ ,  $\bar{\eta}''_t$ , ... are employed for building up the series  $\bar{\zeta}_t$  (cf. (308)). Since such an approach falls under the theory of multi-dimensional stochastical processes, a discussion of this case would be out of place in the present study.

In examining the general methods applied to the BEVERIDGE wheat price index, we have laid stress upon the different type of forecast delivered by the scheme of moving averages as compared with the scheme of hidden periodicities. The approach of moving averages seems particularly useful in forecasting over a short period of time, and attaches directly to current forecast methods, in particular the approach of distributed lag. Having in the previous analysis referred throughout to economic-statistical applications, the present section will be concluded by a few remarks concerning the applicability of the scheme of moving averages in other fields of scientific research.

In periodogram analysis of geophysical data — e.g. records of rainfall, water-levels, temperature, terrestrial magnetism, etc. — it is often difficult to interpret the periods suggested as physical realities. The situation being the same as in economic applications, the claiming of such periodicities has been subjected to severe criticism. A fact especially stressed — see e.g. an excellent critical survey by D. BRUNT (1937) — is that these periodicities can explain but a small, often quite insignificant fraction of the variance in the observational data. In view of this, the lines of research based on multi-dimensional regression analysis seem more promising (see e.g. C. W. B. NORMAND (1932)). Aiming at short time forecasts, the realistic hypothesis underlying these investigations is that the phenomenon considered is causally connected with other phenomena by relations involving distributed lags. The theoretical set-up required having been touched upon in our discussion of economic applications, it is seen that the methods suggested by the theory of moving averages might be used also in these fields of research.

For instance, representing the water-level in a lake by a moving average of the rainfall in surrounding districts, and following the method outlined, we are led to compare the water-level correlogram with the rainfall correlogram as transformed by the hypothetical moving average. By the courtesy of my friend B. BRUNO, who

has taken an interest in my studies in time series analysis, I am in a position to give an illustration of this on the basis of correlograms appearing in a forthcoming paper (B. BRUNO (1938)).

Starting from quarterly observations 1807—1936 of the level of *Lake Väner*; 130 yearly data  $\xi_t$  were obtained by simple averaging. Using formula (13), BRUNO has derived a set of serial coefficients  $r_k(\xi)$  for each of the periods 1807—1936 and 1871—1930. The resulting correlograms are shown in fig. 10.

If a strict periodicity were present in the material, the two correlograms  $r_k(\xi)$  should rise simultaneously to a maximum with

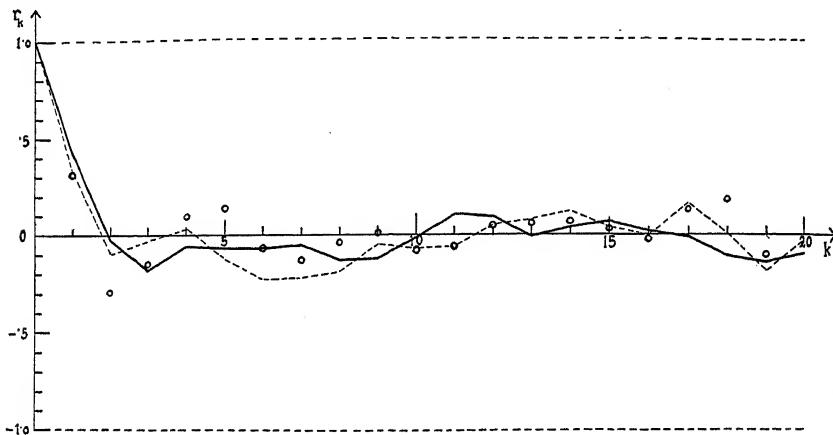


Fig. 10. Correlogram of the level of Lake Väner 1807—1936 (thick line), and 1871—1930 (broken line). Transformed rainfall correlogram (small rings).

abscissa equalling the length of the period. However, the correlograms fluctuating rather independently, no period is suggested in this way. Anyhow, in view of the smallness of the serial coefficients  $r_k(\xi)$  obtained for  $k > 1$ , a hypothetical period cannot be expected to explain a significant part of the variance in the series  $\xi_t$  under analysis.

To my eye, the correlograms suggest more definitely a scheme of moving averages of the simple type  $\{\xi(t)\} = b_0 \{\eta(t)\} + b_1 \{\eta(t-1)\}$ , where  $\{\eta(t)\}$  is purely random. In fact, only the coefficient  $r_1(\xi)$  is rather large in both the correlograms, while — as should be the case if the remaining serial coefficients were chance products — those based on the longer period of observation are, on the whole, lying closer to zero. Taking  $r_1(\xi) = .4$  for the hypothetical value of  $r_1(\xi)$ , we have seen in illustration 2 in section 26 (see. p. 131) that the

corresponding group ( $\zeta$ ) of averages is constituted by the two processes  $\{\eta(t)\} + .5\{\eta(t-1)\}$  and  $.5\{\eta(t)\} + \{\eta(t-1)\}$ .

B. BRUNO has further constructed the correlogram of a series  $\bar{\eta}'_t$  obtained by averaging the yearly rainfall 1867—1936 in four cities in or near the drainage-basin of Lake Väner, viz. Falun, Karlstad, Vänersborg, and Oslo. The correlogram obtained being shown in fig. 11, it is seen that the deviations from zero of  $r_k(\bar{\eta}')$  are rather irregular and small.

Assuming as a first approach the serial coefficients  $r_k(\bar{\eta}')$  to be insignificant, we obtain a closed hypothetical model of the two

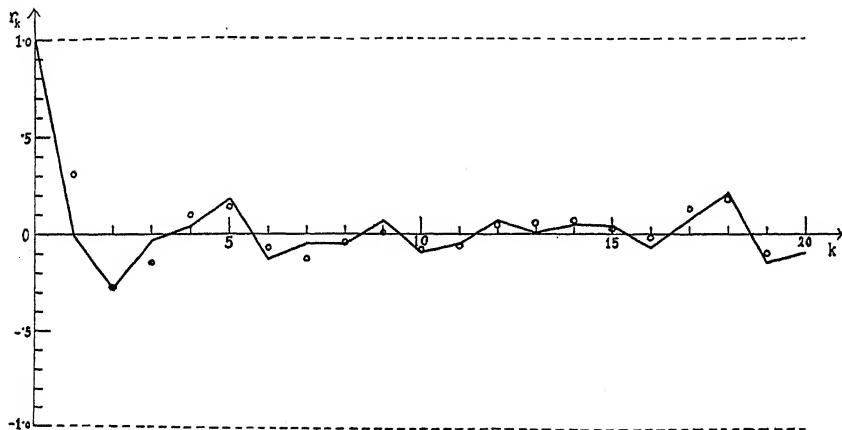


Fig. 11. Correlogram of the rainfall 1867—1936 in the drainage basin of Lake Väner (thick line), and the same correlogram transformed by formula (310) (small rings).

series  $\xi_t$  and  $\bar{\eta}'_t$  by regarding the rainfall series  $\bar{\eta}'_t$  as belonging to a purely random process  $\{\eta'(t)\}$ , and the water-level series  $\xi_t$  as belonging to one of the averages in the group constituted by  $\{\eta'(t)\} + .5\{\eta'(t-1)\}$  and  $.5\{\eta'(t)\} + \{\eta'(t-1)\}$ .

As suggested on p. 170, this simple model might be generalized by cancelling the assumption that  $\{\eta'(t)\}$  be purely random. Letting in such a case the autocorrelation coefficients of  $\{\eta'(t)\}$  be represented by  $r_k(\eta')$ , a short calculation shows that those of  $\{\zeta'(t)\} = b_0\{\eta'(t)\} + b_1\{\eta'(t-1)\}$  will be given by

$$(310) \quad r_k(\zeta') = \frac{(b_0^2 + b_1^2) \cdot r_k(\eta') + b_0 b_1 [r_{k+1}(\eta') + r_{k-1}(\eta')]}{b_0^2 + b_1^2 + 2 b_0 b_1 \cdot r_1(\eta')}$$

(cf. a related formula given by G. U. YULE (1926), Appendix II). Replacing  $r_k(\eta')$  by  $r_k(\bar{\eta}')$ , this formula gives with good approximation

the correlogram of the series  $b_0 \bar{\eta}'_t + b_1 \bar{\eta}'_{t-1}$ . Putting  $b_0 = 1$ , and  $b_1 = .6$ , we have in this manner obtained the transformed correlogram indicated by small rings in fig. 11.<sup>1</sup> For direct comparison with the water-level correlograms, the coefficients  $\bar{r}_k(\bar{\eta}')$  as transformed by (310) have also been plotted in fig. 10. The parallelism with the correlogram based on the period 1871—1930 is rather encouraging — in 10 cases out of 13 the rings and the coefficients  $\bar{r}_k(\bar{\xi})$  vary in the same direction. Observing that the transformation (310) is symmetrical in respect of the coefficients ( $b$ ), we are thus led to examine in detail to what extent the water-level  $\bar{\xi}_t$  may be approximated by a moving average of the rainfall  $\bar{\eta}'_t$  of the simple form  $\bar{\eta}'_t + .6 \bar{\eta}'_{t-1}$  or  $.6 \bar{\eta}'_t + \bar{\eta}'_{t-1}$ .

### 32. Some applications of the scheme of linear autoregression.

Having in Chapter III investigated the process of linear autoregression on the basis of a theory of stochastical difference equations, we shall in the present section give a few applications of this scheme. In doing this, we shall proceed as in the previous section. Choosing an economic time series as our experimental object, we shall first illustrate in detail a general method for determining the parameters when applying a scheme of linear autoregression. The modest purpose being to show how the method works, the tests touched upon in the following discussion will not be applied for examining the significance of the parameters arrived at. In discussing the results, the analysis will instead be focussed on a qualitative comparison with other hypothetical schemes, and with certain related lines of economic-statistical research work. Following up the parallelism with the applications of the scheme of moving averages, this section will be concluded by referring to a few other fields of scientific research which invite an application of the scheme of linear autoregression.

The time series dealt with in the experiments accounted for in the following is the Swedish cost of living index 1830—1913

---

<sup>1</sup> Similarly, (310) gives the serial coefficients of the series  $(\bar{\beta}_t)$  extracted in table 2 in terms of  $\bar{r}_k(\bar{\alpha}_t^{(2)})$  as given on p. 50.

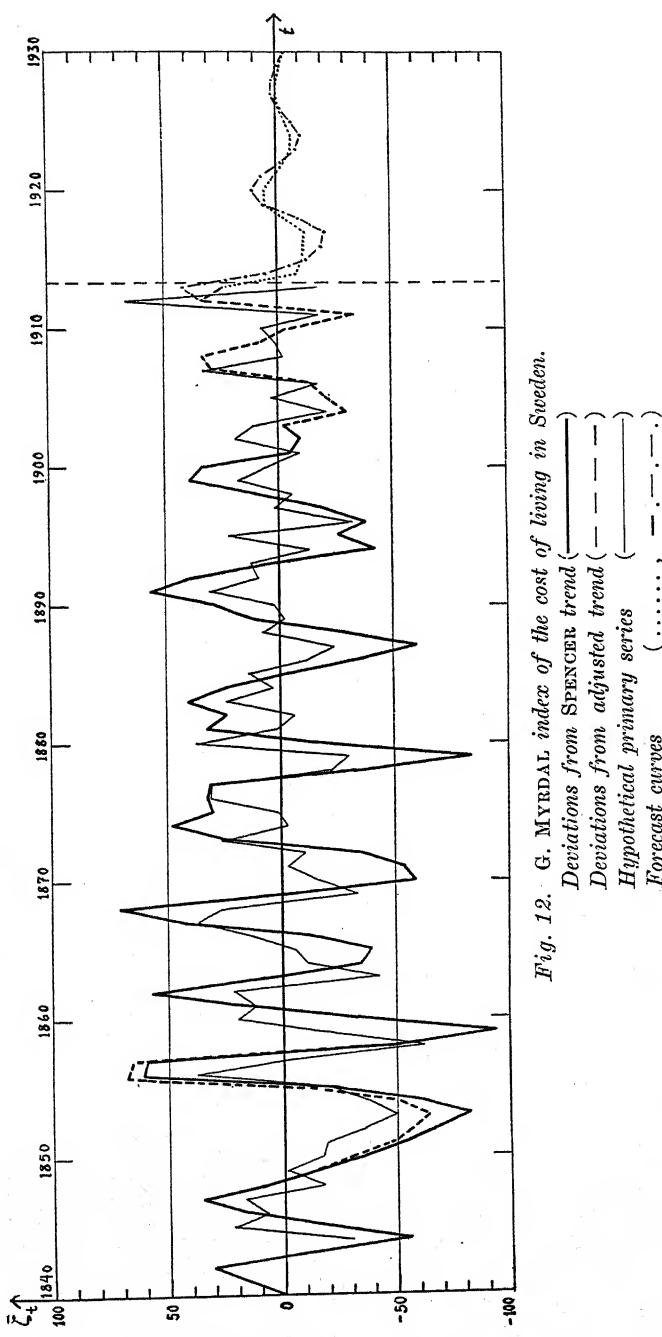


Fig. 12. G. MYRDAL index of the cost of living in Sweden.  
 Deviations from SPENCER trend (— — —)  
 Deviations from adjusted trend (— — —)  
 Hypothetical primary series (— — —)  
 Forecast curves (....., - - - - -.)

compiled by G. MYRDAL ((1933), Table A, budget b). Since the index presents a marked trend, this had to be removed before starting the analysis (cf. p. 146). Having used the 21-term formula of J. SPENCER for this purpose (see e. g. E. T. WHITTAKER and G. ROBINSON (1926), p. 290 f), the deviations from the graduated index are shown in fig. 12 (thick line). The graduation being disturbed by a rapid rise of the index in the years 1851—57, the values obtained for this period were subjected to a slight adjustment (broken line). Another graduation by hand was used for the years 1904—13 not covered by the SPENCER formula (broken line). In order to avoid decimals, the resulting 84 deviations 1840—1913 were multiplied by 10. These index fluctuations, say  $\xi_t$ , constitute our experimental series, and are shown in fig. 12. The numerical values of  $\xi_t$  are given in col. (1) of table 8.

Table 8. Fluctuations in the G. MYRDAL cost of living index (col. 1), and hypothetical primary series  $\bar{n}_t$  (col. 2).

Year	(1)	(2)									
1840	0	.	1860	-33	19·7	1880	-14	36·9	1900	34	5·5
41	16	.	61	21	12·6	81	32	2·1	01	-5	-9·1
42	31	.	62	58	20·4	82	24	-5·9	02	-9	18·8
43	-6	.	63	8	-42·4	83	40	23·3	03	-2	11·2
44	-56	-29·6	64	-35	-11·4	84	24	3·7	04	-30	-21·2
45	-21	21·8	65	-39	-5·2	85	-3	14·5	05	-21	3·6
46	16	6·9	66	-13	12·2	86	-35	-11·9	06	-14	-17·6
47	35	16·9	67	43	36·5	87	-60	-23·6	07	29	31·9
48	9	-18·0	68	71	26·6	88	-26	7·1	08	33	-2·2
49	-13	-1·6	69	-1	-33·8	89	12	-1·9	09	8	'6
1850	-36	-17·0	1870	-59	-16·7	1890	29	2·7	1910	-5	6·1
51	-50	-19·0	71	-54	-3·0	91	56	30·4	11	-34	-17·8
52	-58	-36·4	72	-36	-10·5	92	39	8·8	12	31	65·4
53	-64	-49·0	73	25	26·8	93	2	12·2	13	24	-17·4
54	-48	-38·6	74	48	-3·0	94	-42	-13·3			-198 -47·5
55	-11	-21·1	75	30	1·6	95	-26	21·9			
56	68	38·5	76	32	30·3	96	-38	-32·4			
57	66	-3·7	77	31	30·2	97	-19	1·9			
58	-48	-62·8	78	-33	-22·6	98	11	-6·0			
59	-92	-19·3	79	-84	-30·0	99	39	17·9			

Our series  $\xi_t$  is seen to reflect clearly the changes between economic expansion and contraction. A certain regularity seems to be present in the movement up and down, but the distance between two adjacent maxima is rather inconstant, varying between some 5 and 10 years. The structure of the fluctuations is summed up in the correlogram in fig. 14 obtained from formula (13). The serial coefficients are given in col. (1) of table 9.

The correlogram looks rather like a simple damped oscillation, say  $C \cdot q^k \cdot \cos(\lambda k + \varphi)$ . An inspection of the graph shows that in approximating the correlogram by such a function we would have to take the period  $p = 2\pi/\lambda$  to be about 7 or 8 years, the phase  $\varphi$  to be approximately vanishing, and  $q^7 \sim 1/2$ , the latter relation corresponding to a damping of some 50 % in the duration of one period.

According to the theoretical developments in section 25, a process  $\{\zeta(t)\}$  of linear autoregression as defined by a relation of type

$$(311) \quad \zeta(t) + a_1 \zeta(t-1) + a_2 \zeta(t-2) = \eta(t)$$

will present a correlogram forming a simple damped harmonic. On the other hand, in a scheme of hidden periodicities, each of the harmonic components will give rise to an undamped harmonic in the correlogram. We conclude that such a correlogram cannot approximate the graph of serial coefficients unless at least two harmonics are superposed. However, such a scheme would involve 6 or more parameters, while only two are required in (311). In seeking for a simple hypothetical scheme with correlogram approximating  $r_k$  as shown in fig. 14, we are thus led to try first a process of linear autoregression, and firstly one of the simple type defined by (311).

In a general process (219) of linear autoregression, the linear system (222) with coefficients  $a_1, \dots, a_h$  will deliver the autocorrelation coefficients  $r_1, \dots, r_{h-1}$  required for deriving the following coefficients  $r_h, r_{h+1}, \dots$  etc. from the difference relations (221). In searching for an adequate scheme (311), we are confronted with the inverse problem, viz. to find a set of coefficients  $a_1, \dots, a_h$  giving rise to a hypothetical process (311) with correlogram approximating a prescribed, empirical correlogram. Now, observing that the relations (221) — (223) are linear also in respect of the coefficients ( $a$ ), a convenient starting point for attacking the problem before us would be to replace the coefficients  $r_k$  in (222) and the

last relation (221) — a system whose determinant,  $\Delta(r, n-1)$ , is  $\neq 0$  — by the prescribed values  $r_k$ , and to solve instead for the coefficients  $(a)$ . The system yielding our trial set  $(a)$  thus will read

$$(312) \quad \begin{cases} \bar{r}_1 + a_1 + a_2 \bar{r}_1 + a_3 \bar{r}_2 + \cdots + a_h \bar{r}_{h-1} = 0 \\ \bar{r}_2 + a_1 \bar{r}_1 + a_2 + a_3 \bar{r}_1 + \cdots + a_h \bar{r}_{h-2} = 0 \\ \cdot \quad \cdot \\ \bar{r}_h + a_1 \bar{r}_{h-1} + a_2 \bar{r}_{h-2} + a_3 \bar{r}_{h-3} + \cdots + a_h = 0. \end{cases}$$

If the roots of the equation (34) formed by the resulting coefficients  $(a)$  are lying in the unit circle, these coefficients will define a process (219) of linear autoregression. By construction, the autocorrelation coefficients  $r_1, \dots, r_h$  of this process will coincide with the prescribed values. The following coefficients  $r_{h+1}, r_{h+2}$ , etc. will be obtained from the difference relations (221) formed by the hypothetical coefficients  $(a)$ .

Having derived the correlogram  $r_k$  of the hypothetical process, we are in a position to compare it with the empirical correlogram  $\bar{r}_k$ . If the fit seems satisfactory, we may carry the analysis further on the basis of the coefficients  $(a)$  found, but if the deviations seem too large, an adjustment in the coefficients  $(a)$  is called for.

In analogy with the case of moving averages, we obtain according to (219) a primary series

$$(313) \quad \bar{\eta}_t = \bar{\zeta}_t - \bar{m} + a_1(\bar{\zeta}_{t-1} - \bar{m}) + \cdots + a_h(\bar{\zeta}_{t-h} - \bar{m})$$

corresponding to our set  $(a)$ . In other words, the complete hypothetical model will consist of  $h$  parameters  $(a)$  and the primary series  $\bar{\eta}_t$ ; using these quantities, we can reconstruct the original series  $\bar{\zeta}_t$ . In this case, too, we may look upon the quotient  $\bar{x}^2 = \sum \bar{\eta}_t^2 / \sum \bar{\zeta}_t^2$  as a measure of the efficiency of the analysis — the closer to zero our  $\bar{x}^2$ , the less important is the »unexplained» random component, and the greater the efficiency of the approach.

As mentioned in section 29, the method of G. U. YULE (1927) for determining the coefficients  $(a)$  in the approach (313) is to minimize  $\sum \bar{\eta}_t^2$ . The sum  $\sum \bar{\zeta}_t^2$  being constant, this method will make  $\bar{x}^2$  a minimum, and may thus be said to be of maximum efficiency. Having stated this, let it be pointed out explicitly that the YULE method gives a system for determining the coefficients  $(a)$ , which is equivalent to the system (312) started from above. In fact, minimizing  $\sum \bar{\eta}_t^2$  as defined by (313), we obtain a set of normal equations, where, apart from a constant factor, the coefficients of the

$a_i$ 's will obviously approximate the corresponding serial coefficients in the system (312).

The variance of the process  $\{\eta(t)\}$  defined by our hypothetical set  $(\alpha)$  will be given by formula (220). Of course, this relation holds irrespective of the method used in determining the set  $(\alpha)$ . On the other hand, choosing the system (312) for determining the coefficients  $(\alpha)$ , the resulting primary series  $\bar{\eta}_t$  will evidently satisfy the parallel relation

$$(314) \quad \bar{D}^2(\bar{\eta}_t) \approx (1 + a_1 r_1 + a_2 r_2 + \cdots + a_h r_h) \cdot \bar{D}^2(\zeta_t),$$

where the sign  $\approx$  covers the approximation made in disregarding the first  $h$  terms in the series  $\zeta_t$  for which we cannot calculate corresponding elements  $\bar{\eta}_t$ . In other words, the variance of the primary series  $\bar{\eta}_t$  will approximate the hypothetical value  $D^2(\eta)$ . It must be kept in mind that this will not always be the case if the trial set  $(\alpha)$  is determined otherwise than by (312) (cf. p. 145).

Summing up, the system (312) will give us a set  $a_1, \dots, a_h$  which will minimize the variance of the corresponding residuals  $\bar{\eta}_t$ . Further, the first  $h$  autocorrelation coefficients will coincide with the corresponding serial coefficients. However, the hypothetical correlogram will not always in its whole range yield a good fit to the empirical correlogram. In practice, we must compromise between the two desiderata of obtaining small residuals  $\bar{\eta}_t$  and small deviations between the correlograms, and besides try to satisfy the relation  $D^2(\eta) \sim \bar{D}^2(\bar{\eta}_t)$ . Before discussing this matter, let us see in detail how the method outlined will function when applied to the G. MYRDAL cost of living index.

Forming the system (312) for  $h=2$ , inserting the values  $r_1 = .5216$ ,  $r_2 = -.2240$  given in col. (1) of table 8, and solving for the coefficients  $a_1$  and  $a_2$ , we obtain  $a_1 = -.8771$ ,  $a_2 = .6815$ . The roots of the characteristic equation  $z^2 + a_1 z + a_2 = 0$  being  $.4385 \pm .6994 i$ , and thus less than unity in modulus, we conclude that the relation

$$(315) \quad \zeta(t) - .8771 \zeta(t-1) + .6815 \zeta(t-2) = \eta(t)$$

will define a process of linear autoregression. By construction, the first two autocorrelation coefficients of this process  $\{\zeta(t)\}$  read  $r_1 = \bar{r}_1 = .5216$ ,  $r_2 = \bar{r}_2 = -.2240$ , while the following coefficients will be obtained recurrently from the difference relation  $r_k + .8871 r_{k-1} - .6815 r_{k-2} = 0$ . The resulting correlogram, which is evidently of the form (243), is shown in fig. 14 (thin line).

Table 9. Serial coefficients  $\bar{r}_k$  of the G. MYRDAL cost of living index (col. (1))<sup>1</sup>, and autocorrelation coefficients  $r_k$  belonging to the schemes (317) (col. (2)), and (318) (col. (3)).

$k$	(1)	(2)	(3)	$k$	(1)	(2)	(3)
1	.5216	.5216	.5386	11	-.1533	-.1722	-.2170
2	-.2240	-.2240	-.1460	12	-.2580	-.0218	-.1536
3	-.5811	-.5811	-.5024	13	-.2254	.1065	-.0119
4	-.4626	-.4626	-.5105	14	-.0042	.1311	.1042
5	-.0963	-.0734	-.2320	15	.1883	.0609	.1318
6	.2035	.2749	.1417	16	.1601	-.0338	.0747
7	.3138	.3538	.3458	17	.0723	-.0818	-.0140
8	.2613	.1717	.2902	18	-.0136	-.0630	-.0743
9	.1434	-.0820	.0772	19	-.0067	-.0062	-.0766
10	-.0034	-.2172	-.1321	20	.0342	.0408	-.0827

Comparing with the empirical correlogram, it is seen that the period in the hypothetical correlogram is too short, and that the damping is a little too heavy. According to section 6, the damping factor equals  $\sqrt{a_2}$ , while the period is given by  $p = 2\pi/\lambda$ , where  $\cos \lambda = -a_1/2\sqrt{a_2}$ . Thus, an increase in  $a_2$  will bring on a slighter damping. Further, reducing  $\lambda$  we obtain a longer period. However, as pointed out in the previous discussion, we cannot conclude without further evidence that it will be possible to improve the fit — the coefficients  $a_1$  and  $a_2$  determine also the constant factor and the phase of the damped harmonic, and it might happen that an adjustment in  $a_1$  and  $a_2$  would cause such a change, e. g. in the phase, that the total result of the adjustment would be a poorer fit. Proceeding with the illustration, we shall next examine the total effect of an adjustment.

In the correlogram of the approach (315), the period is found to be 6.22 years, while a good fit would require a period not below 7 years. Reducing the damping by increasing  $a_2$  from .6815 to .77, a short calculation will verify that  $a_1 = 1.10$  will give a period  $p = 7.03$  years. Now, let us examine the approach

$$(316) \quad \zeta(t) - 1.10 \zeta(t-1) + .77 \zeta(t-2) = \eta(t).$$

<sup>1</sup> In reading the proofs, a slight error was discovered in the serial coefficients: they should all have been multiplied by a factor amounting to about 1.005.

The correlogram of the process  $\{\zeta(t)\}$  thus defined has been calculated from (221) and (222), and is shown in fig. 14 (broken line). Up to  $r_8$  and  $r_9$ , the hypothetical correlogram seems to fit rather well. Beyond this point, the fit is less satisfactory, partly because the graph of serial coefficients presents a slow descent to the minimum in  $k \sim 12.5$ , and a rapid rise to the next maximum. This skewness will be recurred to later.

A clear view of the adjustment will be obtained by calculating the roots of the characteristic equations of (315) and (316). In fig. 13, these roots are indicated by small rings. In drawing the conjugate roots nearer to the periphery of the unit circle, the damping has been reduced, while the reduction in the angle  $\lambda$  has elongated the period.

Each of the approaches (315) and (316) gives rise to a primary series  $\bar{\eta}_t$ , and we know from a previous remark of general scope that the improvement in the fit of the hypothetical correlogram is obtained at the expense of an increase in the variance of the series  $\bar{\eta}_t$ . In addition, the approach (315) being based on a system of type (312), the relation (314) will hold.

Thus, paying regard to (220), and inserting  $a_1 = -0.8771$ ,  $a_2 = 0.6815$ ,  $r_1 = 0.5216$ ,  $r_2 = -0.2240$ , we obtain in this case  $\bar{D}^2(\bar{\eta}_t) \sim D^2(\eta) = 390 D^2(\zeta)$ . As to the residuals  $\bar{\eta}_t$  derived from the approach (316), the development

$$(317) \quad \bar{D}^2(\bar{\eta}_t) = \bar{D}^2(\bar{\zeta}_t + a_1 \bar{\zeta}_{t-1} + a_2 \bar{\zeta}_{t-2}) \sim \\ (1 + a_1^2 + a_2^2 + 2a_1 \bar{r}_1 + 2a_2 \bar{r}_2 + 2a_1 a_2 \bar{r}_1) \cdot \bar{D}^2(\bar{\zeta}_t)$$

will not reduce to (314). Inserting  $a_1 = -1.10$ ,  $a_2 = .77$ ,  $\bar{r}_1 = 0.5216$ ,  $\bar{r}_2 = -0.2240$ , we find in this case  $\bar{D}^2(\bar{\eta}_t) \approx 427 \bar{D}^2(\bar{\zeta}_t)$ . In full agreement with the general theory, the adjustment in the coefficients ( $a$ )

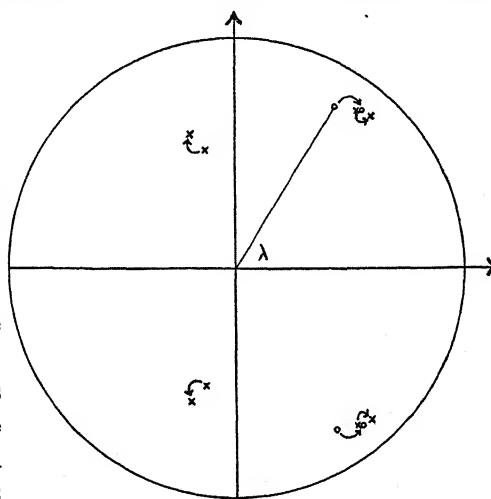


Fig. 13. Adjustment paths in the approaches (315) (rings), and (318) (crosses).

has reduced the efficiency of the approach. As mentioned in connexion with (314), the hypothetical variance  $D^2(\eta)$  will also be affected by the adjustment. Generally speaking, nothing compels  $D^2(\eta)$  to follow the variation in  $\bar{D}^2(\bar{\eta}_d)$ . Actually, in the present case  $D^2(\eta)$  varies contrarily to  $\bar{D}^2(\bar{\eta}_d)$  — the first two autocorrelation coefficients in the scheme (316) being  $r_1 = .6215$ ,  $r_2 = -.0864$  (cf. fig. 14), it is readily verified by inserting these values together with  $a_1 = -1.10$ ,  $a_2 = .77$  in formula (220) that the variance in question will be given by  $D^2(\eta) = .250 D^2(\zeta)$ .

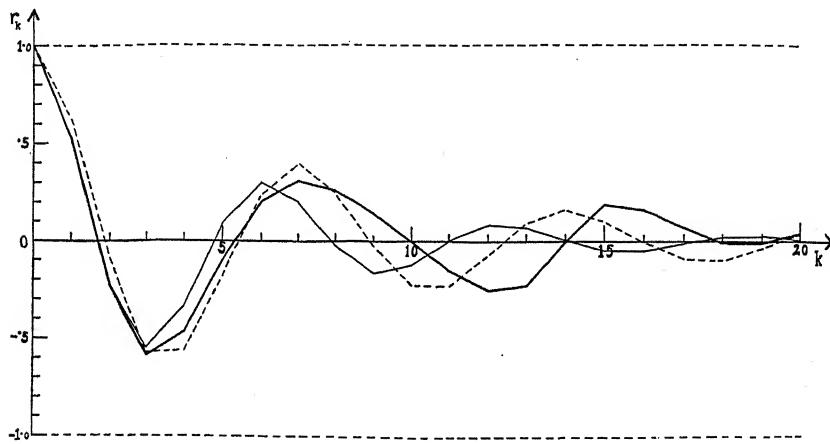


Fig. 14. Correlogram of the G. MYRDAL cost of living index (thick line), and hypothetical correlogram corresponding to formula (315) (thin line), and formula (316) (broken line).

Summing up the comparison between the approaches (315) and (316), the hypothetical correlogram fits better in the latter case, but the variance  $\bar{D}^2(\bar{\eta}_d)$  is smaller in the former case and coincides with the hypothetical variance  $D^2(\eta)$  — in the approach (316) the difference between these variances is rather large. All in all, neither of the schemes seems adequate. In view of other experiments with approaches of the simple type (311), it seems as if we cannot find a satisfactory approach without taking into account more distant elements  $\xi_{t-s}$ ,  $\xi_{t-4}$ , etc.

Of course, having different desiderata to comply with in applying a scheme of linear autoregression, we could agree on which weights to attach to them, and then take them into consideration simultaneously. In point of principle, it would then be possible to find

out a set ( $\alpha$ ) forming the best compromise in the sense agreed. However, judging from certain experiments of this kind, what might be gained in this way seems not worth the extensive computations involved. One way or another, the results arrived at in these experiments merit no recital. Accordingly, proceeding to an account of certain experiments with a scheme (219) involving four parameters ( $\alpha$ ), we shall follow the same line as before.

Taking  $h = 4$ , and inserting in the system (312) the serial coefficients  $\bar{r}_k$  given in table 9, we arrive at the following approach

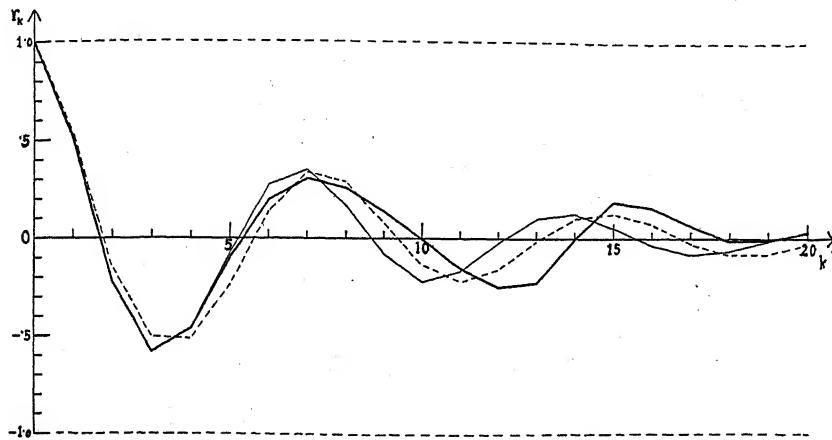


Fig. 15. Correlogram of the G. MYRDAL cost of living index (thick line), and hypothetical correlogram corresponding to formula (318) (thin line), and formula (319) (broken line).

$$(318) \quad \zeta(t) - .8100 \zeta(t-1) + .7452 \zeta(t-2) - .0987 \zeta(t-3) + \\ + .2101 \zeta(t-4) = \eta(t).$$

The autocorrelation coefficients of the process  $\{\zeta(t)\}$  thus defined have been calculated from (221) and (222). The values found are given in col. (2) of table 9, and plotted in fig. 15 (thin line).

It is seen that the approaches (316) and (318) give rise to almost coincident correlograms, only that here we have by construction  $r_k = \bar{r}_k$  for  $k \leq 4$ . As shown in fig. 13, this conformity is reflected in the characteristic equations — in the present case two of the four roots are lying in the close neighbourhood of the roots belonging to the approach (316). The numerical values found for the roots in the approach (318) read  $.5385 \pm .6814 i$ ,  $-.1335 \pm .5106 i$ .

In adjusting the approach (318), we are at liberty to move the roots of the characteristic equation in any directions, keeping in mind that complex roots must be conjugate. The dominant component in the correlogram evidently corresponds to the roots  $5385 \pm 6814 i$ , say  $\rho \cdot e^{\pm i\lambda}$ . The period of this component is nearly 7 years, while the empirical correlogram suggests a somewhat longer period. Now, reducing the angle  $\lambda$  so as to obtain a period equaling 7.5 years, the resulting set of coefficients ( $a$ ) gave rise to a correlogram with reduced amplitudes. Neutralizing this effect by reducing the damping by means of a slight move towards the periphery of the unit circle, it was found adequate to perform a simultaneous move in the other two roots. As a matter of fact, the deviations between  $r_k$  and  $\bar{r}_k$  for  $k = 1 - 4$  caused by the adjustments mentioned were found to be substantially reduced by moving the root  $-1335 + 5106 i$  in a direction nearly opposite to the adjustment in the root  $5385 + 6814 i$ .

Having thus arrived at the roots  $5888 \pm 6540 i$ ,  $-20 \pm 58 i$ , the paths followed are indicated in fig. 13. As is readily verified, the adjusted roots belong to the approach

$$(319) \quad \zeta(t) - 7776 \zeta(t-1) + 6797 \zeta(t-2) - 1342 \zeta(t-3) + \\ + 2914 \zeta(t-4) = \eta(t).$$

The autocorrelation coefficients of this scheme as derived from the system (222) and the relations (221) are given in col. (3) of table 9, and plotted in fig. 15 (broken line). The fit to the empirical correlogram is not very close, but the general shape of the hypothetical correlogram is rather satisfactory. Comparing with the approach (316), which involves only two parameters, it is seen that the improvement bears chiefly upon the period and the coefficient  $r_1$ .

Using formula (314), and paying regard to the identities  $r_k = \bar{r}_k$ ,  $k \leq 4$ , the approach (318) gives  $D^2(\eta) = 371 D^2(\zeta)$ ,  $\bar{D}^2(\eta) = 371 \bar{D}^2(\zeta)$ . Comparing with the simple scheme (315), it is seen that the reduction of the factor in  $D^2(\eta)$  and  $\bar{D}^2(\eta)$  amounts to only 0.019. In other words, the introduction of two more parameters has brought on but a slight increase in the efficiency of the approach. However, comparing the adjusted schemes, we find a definite improvement. Proceeding as in (317), we find in the first place that the residuals derived from (319) will satisfy the relation  $\bar{D}^2(\bar{\eta}) \approx 381 \bar{D}^2(\bar{\zeta})$  (working directly on the series  $\zeta_t$  and  $\bar{\eta}_t$  given in table 8, we find

$\bar{D}^2(\bar{\eta}_t) = .385 \bar{D}^2(\bar{\zeta}_t)$ . The adjustment has thus reduced the efficiency of the approach but very slightly. On the other hand, applying formula (220), we find that the parallel hypothetical relation reads  $D^2(\eta) = .401 D^2(\zeta)$ . Contrary to the situation in the approach (316), it is seen that  $\bar{D}^2(\bar{\eta}_t)$  in the present case approximates  $D^2(\eta)$  even after the adjustment.

Perhaps it would be possible to find an adjustment improving the approach (319). However, in view of the above figures, not much can be gained by a continued adjusting. Nor does it seem

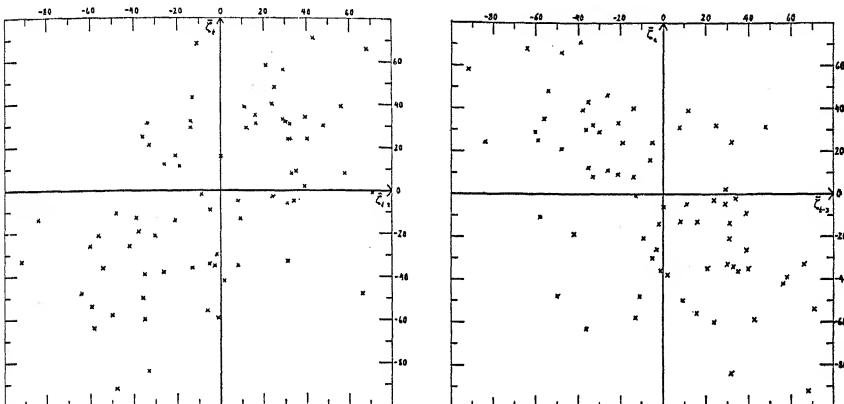


Fig. 16. G. MYRDAL cost of living index  $\xi$ . The scatters  $(\xi_t, \xi_{t-1})$  (left), and  $(\xi_t, \xi_{t-3})$  (right).

as if a real improvement could be secured by enlarging our set (a). Be that as it may, the above examples are sufficient for our purpose of illustrating a general method for deriving a trial set of coefficients (a) in applying a process of linear autoregression, and for performing adjustments in the trial set found.

Proceeding to a discussion of the approach (319), the same general viewpoints present themselves as in the case of moving averages. Accordingly, referring to the remarks in the previous section (cf. p. 163), we need draw attention only to a few circumstances which are peculiar to the scheme of linear autoregression.

As indicated by the term proposed, the hypothesis of linear autoregression implies a linear regression upon  $\xi_t$  of each of the preceding elements  $\xi_{t-1}$ ,  $\xi_{t-2}$ , etc. This circumstance having already been employed by G. U. YULE (1927) for testing an approach of this kind (cf. section 29, p. 141), we show in fig. 16 the scatters

$(\xi_t, \xi_{t-1})$  and  $(\xi_t, \xi_{t-3})$  of the fluctuations in the MYRDAL index. The deviations from linearity in the connexion between the variables do not seem disturbing.

As in the case of moving averages, different tests of a scheme of linear autoregression may be based on the hypothetical randomness of the primary series  $\bar{\eta}_t$ . Such tests may, for example, be focussed upon the scatters  $(\bar{\eta}_t, \bar{\eta}_{t-1})$ . The scatter  $(\bar{\eta}_t, \bar{\eta}_{t-1})$  formed by the residuals  $\bar{\eta}_t$  given in table 8 is shown in fig. 17.

The above mentioned skewness in the correlogram of the G. MYRDAL index (see fig. 14) gives an interesting illustration of the difficulties of testing time series schemes. Is it permitted to look upon the deviation from the hypothetical correlogram as produced by pure chance? An examination of this question must pay due regard to the interdependence of the serial coefficients in a sample series. For instance, a chance deviation in  $k=13$  would perhaps be most often attended by such deviations in the neighbouring coefficients that the total picture would present a skew oscillation. The question of how much weight to attach to deviations of this and similar kinds seems extremely intricate. Perhaps nothing better can be done than to compute a large number of model series correlograms, and then to compare the deviations from the hypothetical curve.

As pointed out in section 24, an approach of linear autoregression will give rise to a forecast curve which forms a damped oscillation of type (33), and satisfies the same linear difference equation as the autocorrelation coefficients. Letting as before  $F_t[\zeta(t+k)]$  represent a forecast over  $k$  time units, formula (214) shows that  $F_t[\zeta(t+k)]$  may be conveniently computed recurrently. For instance, considering the approach (319), we get  $F_t[\zeta(t+1)] = (1 + a_1 + a_2 + a_3 + a_4) \cdot \bar{m} - a_1 \xi_t - a_2 \xi_{t-1} - a_3 \xi_{t-2} - a_4 \xi_{t-3}$ ,  $F_t[\zeta(t+2)] = (1 + a_1 + a_2 + a_3 + a_4) \cdot \bar{m} - a_1 \cdot F_t[\zeta(t+1)] - a_2 \xi_t - a_3 \xi_{t-1} - a_4 \xi_{t-2}$ , etc. In particular, inserting for  $t=1912$  the values of  $\xi_t$  given in col (1) of table 8, we find  $F_{1912}[\zeta(1913)] = 41.4$ ,  $F_{1912}[\zeta(1914)] = 5.2$ , etc. The forecast curve  $F_{1912}[\zeta(1912+k)]$  is shown in fig. 12 (dashes and dots). A check of the first forecast is obtained by deducting

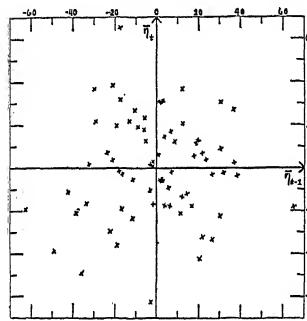


Fig. 17. Scatter  $(\bar{\eta}_t, \bar{\eta}_{t-1})$  of a hypothetical primary series of G. MYRDAL's index of the cost of living.

the hypothetical residual (»random shock»)  $\bar{\eta}_{1913} = -17.4$ ; the result is 24, which equals  $\bar{\zeta}_{1913}$ . Fig. 12 shows also the forecast curve  $F_{1913}[\zeta(1913 + k)]$ , (dotted line).

The two forecast curves in fig. 12 yield a good illustration of the prognosis situation in an approach of linear autoregression. Firstly, while a forecast  $F_t[\zeta(t+k)]$  is often rather efficient for small  $k$ -values, the efficiency vanishes asymptotically as  $k$  increases (cf. p. 165). Further, as soon as we are in a position to take a new observation  $\zeta_{t+1}$  into consideration when forming the prognosis, the forecast curve is often substantially modified; how much, will depend on the residual  $\bar{\eta}_{t+1} = \bar{\zeta}_{t+1} - F_t[\zeta(t+1)]$ . — Summing up, it is the short forecasts that are efficient. In this respect, we meet the same situation as in the scheme of moving averages, and the same contrast to the scheme of hidden periodicities (cf. p. 168). On the other hand, under special circumstances the oscillations in a scheme of linear regression are nearly functional, viz. nearly strictly periodic — as remarked in discussing the sinusoidal limit theorem of E. SLUTSKY (cf. p. 120), processes of hidden periodicities can be obtained as limit cases of the schemes of linear autoregression.

As pointed out in section 21, the scheme of linear autoregression constitutes the proper starting point when studying oscillatory mechanisms which are subjected to random impulses. A typical approach of this kind is formed by the *complete systems* as dealt with in several recent economic studies (see e. g. R. FRISCH (1933), J. TINBERGEN (1937)). A simple example of a complete system is given by

$$(320) \quad \begin{cases} \xi(t) = c_1 \zeta(t-1) + \eta'(t), \\ \zeta(t) = d_0 \xi(t) + d_1 \xi(t-1) + \eta''(t). \end{cases}$$

For instance, as a first approximation we may take the production volume  $\xi(t)$  of a commodity to be linearly correlated with the price  $\zeta(t-1)$ , and the price  $\zeta(t)$  to be linearly built up by the production volumes  $\xi(t)$  and  $\xi(t-1)$ .

In this connexion, the pertinent thing is that a complete system may be reduced to a single relation involving but one of the fundamental variables  $\xi$ ,  $\zeta$ , etc. Considering e. g. the simple case (320), we get at once

$$(321) \quad \zeta(t) + a_1 \zeta(t-1) + a_2 \zeta(t-2) = \eta(t),$$

where  $a_1$  and  $a_2$  are constants, and  $\eta$  is linear in the variables  $\eta'$  and  $\eta''$ .

Of course, in order to study a complete system in detail, we must consider stochastical processes in several dimensions, a generalization not in the program of the present study. However, it is evident that theorem 9 applies directly to the reduced relations exemplified by (321). Thus, under general conditions concerning the variables  $\zeta(t)$  and the constants  $(a)$ , the variables  $\zeta(t)$  form a stationary process. In point of principle, we are in a position to investigate the properties of this process  $\{\zeta(t)\}$ . Considering e.g. the autocorrelation coefficients, it follows without difficulty that these will satisfy a linear difference equation with a right member.

It is seen that the theory of linear autoregression as developed in sections 24 and 25 covers the case when the complete system reduces to a relation with a purely random right member. We have seen that this analysis has given certain results which cannot be reached by functional methods. For instance, if the left member of the relation (321) is characterized by an intrinsic damped oscillation, and if the damping is too heavy, the tendency to periodicity cannot be distinguished in the mechanism as subjected to random impulses. Another example of this is given by the approach (319), which presents two intrinsic periods. Having already mentioned that one of these equals 7·5 years, a short calculation will verify that the other period is 3·44 years. Now, as shown in fig. 13, the root corresponding to the latter period is lying rather near the centre of the unit circle, which implies that the damping is rather heavy. Thus, even if the period 3·44 is quantitatively reliable and significant — which seems to me rather doubtful — we cannot conclude without further analysis that this period can be found by a periodogram construction or similar methods. Be that as it may, as pointed out before I attach no importance to the quantitative significance of the above analysis of the MYRDAL index.

Starting from explicit assumptions about the relations between the different variables in an economic system, E. LUNDBERG (1937) has examined how the system will develop from hypothetical initial conditions. Since several of the relations assumed are non-linear, his approach may be looked upon as a generalization of the linear systems as exemplified by (320). Now, the analysis of E. LUNDBERG shows that the variables will often present tendencies to diverge

as an economic expansion goes on, tendencies causing tensions which make the economic system instable. Of course, in point of principle it would be possible to apply stochastical methods even in this approach. However, it seems extremely difficult to obtain in this way a non-evolutive scheme for the system considered. Anyhow, in view of the investigations of E. LUNDBERG, the linear approaches earlier discussed seem far from sufficient for giving an adequate hypothetical model for studying the economic cycles in detail. Looking upon the approach (319) from this viewpoint, it might be said that even if this scheme does correctly sum up the main features of the index examined, the interpretation in terms of oscillatory mechanisms which is suggested by such a simple approach cannot possibly be completely realistic.

Having in the previous section mentioned certain geophysical phenomena which invite to studies on the basis of the scheme of moving averages, this section will be terminated by a few suggestions about the wide applicability of the scheme of linear autoregression.

As surveyed in section 29, G. U. YULE (1927) introduces the concept of autoregression in studying the 11 year wave in the sunspot numbers. Further, since the criterion of J. BARTELS (1935) (cf. p. 26) suggests that certain waves in terrestrial magnetism are »quasi-persistent», the construction of this criterion makes us expect that in these cases an autoregression approach will be fruitful. Of course, here the scheme of linear autoregression suggests itself also *a priori*. For instance, let us consider the 27 day wave, which is due to the sunspot intensity and the rotation of sun. The duration of a sunspot often being rather long, the sunspot intensity as observed in a time point  $t$ , say  $\xi_t$ , must be positively correlated with the intensity 27 days earlier. Having stated this, it seems plausible that a correlogram of terrestrial magnetism will present an oscillation with a period of about 27 days. We must also expect that the oscillation in the empirical correlogram will be damped, and that the degree of damping will depend on the average duration of a sunspot.

It is not difficult to point out other instances where the scheme of linear autoregression seems plausible on theoretical grounds. Even the simple scheme (237) involving but one parameter might often prove useful, at least as a first approximation. For instance, we have already found that the scheme (284) presents a correlogram which fits rather well to the correlogram of air pressure examined

by SIR G. WALKER (1931). In this case, the autoregression obviously may be interpreted as an effect of inertia — generally speaking, we may always expect an autoregression of type (237), and with a positive constant  $p$ , when dealing with phenomena characterized by irreversibility. According to formula (239), the expectance corresponding to a long period is in such cases larger than in a purely random series. In other words, there will be a tendency to spurious periodicity, a tendency of the observational time series to present long waves the lengths of which are varying and without physical significance.

Finally, it is evident that the theory of autocorrelation may be applied to the functional transform (21) used by N. WIENER (1930) in the theory of light. Following up this idea, we are led to interpreting the transmission of light as a stationary process. With suitable arrangements about the dispersion of the process, the expression (21) would then correspond to the autocorrelation coefficients, while the function  $S(\lambda)$  as given by (23) would reduce to the generating function of the autocorrelation coefficients (cf. section 17). In this approach, the continuous parts of the spectrum of light would correspond to the continuity intervals of the generating function. Perhaps the scheme of linear autoregression might serve as a starting point for investigations on these lines, for according to theorem 11 this scheme presents a continuous generating function, and, as pointed out in section 25 (see p. 120), we may obtain any scheme of hidden periodicities of type (39) as a limit case.

## APPENDIX A.

### On the $\omega^2$ -method for testing goodness of fit.

Let  $\bar{F}(u)$  denote an empirical distribution function, and  $F(u)$  a hypothetical distribution function to be tested.

In the  $\omega^2$ -method for testing the goodness of fit, the integral

$$(322) \quad \omega^2 = \int_{-\infty}^{\infty} [\bar{F}(u) - F(u)]^2 du$$

is evaluated and compared with its expectation.<sup>1</sup> This is given by

$$(323) \quad E[\omega^2] = \frac{1}{n} \cdot \int_{-\infty}^{\infty} F(u)[1 - F(u)] du = \frac{g}{2n},$$

where  $n$  is the number of observations in the empirical material, and  $g$  is the mean difference corresponding to the distribution function  $F(u)$  (see H. CRAMÉR (1928), p. 145, and, as to the second formula, H. WOLD (1935), p. 48).

The  $\chi^2$ -method considers the frequency distribution in a set of intervals. Let

$$(324) \quad u_0 = -\infty, u_1, u_2, \dots, u_k = +\infty$$

stand for the end points of the intervals considered, and denote the frequency in the interval  $(u_i, u_{i-1})$  by

$$(325) \quad \bar{f}_i = \bar{F}(u_i) - \bar{F}(u_{i-1}); \quad f_i = F(u_i) - F(u_{i-1}).$$

In these notations the  $\chi^2$ -test is given by

$$\chi^2 = \frac{n}{k-1} \cdot \sum_{i=1}^k (\bar{f}_i - f_i)^2 / f_i,$$

and  $E[\chi^2] = 1$ .

<sup>1</sup> The N. SMIRNOFF (1936) modification of the  $\omega^2$ -test, viz.  $\omega^2 = \int_{-\infty}^{\infty} [\bar{F}(u) - F(u)]^2 dF(u)$ , has the beautiful property that the probability distribution of the test is independent of  $F(u)$ .

As is well known, the weak point in the  $\chi^2$ -method lies in the arbitrariness implied in the required grouping of frequencies. This defect is especially fatal in the tails of the distribution, where the frequencies  $f_i$  are small, and the terms in  $\chi^2$  are very susceptible to changes in the grouping. For the same reason the  $\chi^2$ -method is always inapplicable if the number of observations is small.

The above mentioned defect does not adhere to the  $w^2$ -method, as the frequency grouping is not needed here. However, in respect to the  $w^2$ -method it might be maintained that the numerical evaluation of the integral (322) is toilsome. Further, if the statistical material is from the start given in the form of grouped frequencies  $\bar{f}_i$ , the function  $\bar{F}(u)$  required for the test will be known only in the end points  $u_i$  of the class intervals. In the following slight modification of the  $w^2$ -test, both of these obstacles are removed.

The modified test is defined by

$$(326) \quad w^2 = \sum_i [\bar{F}(u_i) - F(u_i)]^2,$$

where the points  $u_i$  may be taken equidistant if the material is ungrouped. If the material is given as grouped frequencies (325), the  $u_i$ 's in (326) must, of course, be chosen among the  $u_i$ 's in (324).

It is seen that the modified test (326) is simple to compute. Further, the arbitrariness in (326) as to the points  $u_i$  is unimportant. In fact, they may be taken as close as desired, and have no systematic influence upon  $w^2$ .

The expectation of  $w^2$  is easily derived:

$$(327) \quad E[w^2] = \sum_i E[\bar{F}(u_i) - F(u_i)]^2 = \frac{1}{n} \cdot \sum_i F(u_i)[1 - F(u_i)],$$

an expression readily computed in practice. In case the points  $u_i$  are equidistant, say  $u_i = u + i \cdot h$ , formula (327) gives

$$(328) \quad E[w^2] = \frac{1}{h \cdot n} \cdot \int_{-\infty}^{\infty} F(u)[1 - F(u)] du + \frac{1}{n} \cdot R,$$

where  $R$  is the remainder in the EULER-MACLAURIN formula applied. Neglecting  $R/n$ , formula (328) may be used alternatively with (327).

*Illustration.* During the years 1920—1934, a rural fire insurance company in Sweden had the following yearly claim ratios, as expressed in per mille: '210, '1042, '261, '467, '708, '412, '455, '853, '641, '289, '669, '071, '303, '097, '1039. The distri-

bution function  $\bar{F}(u)$  obtained from these figures is shown in the diagram below together with a PEARSON Type III distribution function fitted to the data, *viz.*

$$F(u) = U_0 \cdot \int_0^u x^p \cdot e^{-\gamma x} dx \text{ for } u \geq 0, \quad F(u) = 0 \quad \text{for } u \leq 0,$$

with  $p = 1.725$  and  $\gamma = 5.437$ .

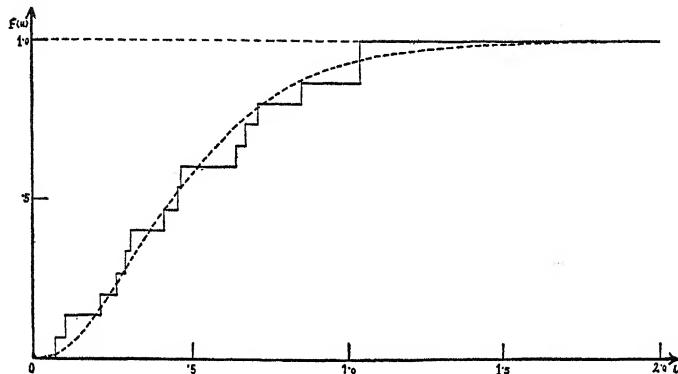


Fig. 18. Hypothetical distribution function  $F(u)$  (broken line) fitted to an empirical distribution function  $\bar{F}(u)$  (unbroken line).

For testing the goodness of fit, formulae (326) and (328) were used. It should be observed that the material under investigation is too small for the application of the  $\chi^2$ -test. Observing that  $n = 15$ , the simple computations may be followed in the table below.

Table 10. Application of the  $\omega^2$ -test.

$u$	$\bar{F}(u)$	$F(u)$	$1 - F$	$ \bar{F} - F $	$u$	$\bar{F}(u)$	$F(u)$	$1 - F$	$ \bar{F} - F $
.05	.00	.01	.99	.01	.85	.80	.87	.13	.07
.15	.13	.07	.93	.06	.95	.87	.91	.09	.04
.25	.20	.21	.79	.01	1.05	1.00	.94	.06	.06
.35	.40	.38	.62	.02	1.15	1.00	.96	.04	.04
.45	.47	.52	.48	.05	1.25	1.00	.97	.03	.03
.55	.60	.64	.36	.04	1.35	1.00	.98	.02	.02
.65	.67	.74	.26	.07	1.45	1.00	.99	.01	.01
.75	.80	.81	.19	.01	1.55	1.00	.99	.01	.01

There only remains to compute  $\omega^2 = \sum_{15} |\bar{F} - F|^2$  and its expectation  $E[\omega^2] = \sum F(1 - F)$ . As is easily verified,  $\omega^2 = .0265$ , and  $E[\omega^2] = .1107$ , which indicates a very nice fit.

It should be observed that in the above illustration the reduction of  $E[\omega^2]$  in accordance with the degree of freedom in  $n$  has not been taken into consideration.

## APPENDIX B.

### On the quantitative significance of correlation coefficients.

It might appear paradoxical that correlation analysis, while of fundamental importance in many sciences, is in others severely criticized and denied quantitative significance. Let us exemplify this contrast. In genetics, the fundamental facts are often expressed in terms of correlation coefficients; the correlation between the statures of father and son is a classical example. The correlation coefficient is here in possession of an undisputed quantitative significance. On the other hand, it is frequently argued that we are concerned with a different, non-quantitative kind of correlation coefficient as often as the coefficients have been obtained from time series or spatial series.

In the discussion of the significance of correlation coefficients, a chief argument is based on the general experience that time series and spatial series often give rise to nonsense correlation, or correlation coefficients from which it is difficult or impossible to draw quantitative conclusions. To a certain extent, this experience has been analysed theoretically. Thus, the effect of trends in time series has been studied in detail. In the terminology of the present volume, the trend problem in question may be outlined as follows.

Let  $\xi_t, \xi_{t+1}, \dots, \xi_{t+n}$  and  $\eta_t, \eta_{t+1}, \dots, \eta_{t+n}$  represent two time series, and let two corresponding hypothetical sets of random variables be given by  $x(t) + \xi(t)$  and  $y(t) + \eta(t)$ , where  $x(t)$  and  $y(t)$  are functions, while  $\xi(t)$  and  $\eta(t)$  are random variables (cf. (39)). When studying the relations between the series  $\xi_t$  and  $\eta_t$ , it may be that the interest is concentrated upon the variables  $\xi(t)$  and  $\eta(t)$ , thus upon the coefficient  $r[\xi(t), \eta(t)]$  instead of  $r[x(t) + \xi(t), y(t) + \eta(t)]$ . As a rule, this happens when  $x(t)$  and  $y(t)$  are trend functions, e.g. parabolas, exponential or logistic curves; the relevant thing is then

to study the correlation between the trend deviations  $\xi(t)$  and  $\eta(t)$ . Otherwise expressed, an interdependence between the two series is produced by time itself, i. e. by the common variable  $t$  in the two functions  $x(t)$  and  $y(t)$ . Generally speaking, it depends on the nature of the phenomena described by the time series whether or not this interdependence should be removed, i. e. whether  $r[x(t) + \xi(t), y(t) + \eta(t)]$  or  $r[\xi(t), \eta(t)]$  should be preferred as a measure of the correlation between the two time series considered. This remark brings into relief the necessity both of formulating the problems strictly when studying time series, and of doing this in terms of explicit hypothetical models.

A second introductory reference is given to a note of V. v. SZELISKI (1934). In a time series  $\xi_t, \dots, \xi_{t+n}$  which consists of a sequence of frequencies  $\xi_{t+k}$  it may occur, v. SZELISKI says, that the recording has been disturbed by *alternating errors*, i. e. such mistakes that one or more elements are included in  $\xi_{t+k}$  instead of in  $\xi_{t+k+1}$ . Considering the correlation between two time series  $\xi_t$  and  $\eta_t$ , such alternating errors tend to diminish the correlation coefficient. Moreover, if the time intervals are enlarged — e. g. so that the frequencies  $\xi_t + \xi_{t+1}, \xi_{t+2} + \xi_{t+3}$ , etc. are recorded instead of  $\xi_t, \xi_{t+1}$ , etc. — the effect of the alternating errors will decrease.

The above references to the trend problem and to the alternating errors show that correlation analysis is more intricate in time series than in such classical applications as the genetic example. The purpose of this appendix (cf. also p. 109 f. and my preliminary note (1936)) is to draw attention to a circumstance which complicates the situation still further. The point in question is based on an elementary distinction, viz. between what I have proposed to term *definite* and *modifiable* statistical units. It will be found that a correlation index is unconditionally quantitative only when referring to a definite unit. On the other hand, correlation indices which refer to modifiable units are not directly commensurable.

In the case of modifiable units, the statistical analysis requires more refined methods. We advance further that modifiable units are very common in the cases of time series and spatial series. The subject being wide and uninvestigated, we shall not aim at generality in our analysis, but instead examine in some detail a typical case. Since the distinction between definite and modifiable units is quite elementary, we must begin by analyzing some fundamental statistical concepts.

Statistical data always have reference to a statistical population formed by statistical units or individuals; elementary, unprepared statistical data always refer to some property of the units, and describe quantitatively — by class-numbers or by measure-numbers — the individual units in respect of the property considered. Having stated this, a statistical unit will be termed *definite* if its delimitation is by nature unique. Otherwise, i. e. if any arbitrariness attaches to its size, the unit will be termed *modifiable*. For the sake of concreteness, let us at once exemplify the distinction made.

Recurring to the correlation between the statures of father and son, a specific correlation coefficient may always be taken to refer to a statistical population formed by a group of families. The two properties of the statistical unit, the family, which are correlated, are the stature of the family father and the stature of, say, his eldest son. In this example, the statistical unit is definite.

Secondly, let us examine the claim ratios considered in Appendix A. In this case, the statistical units are constituted by a certain rural district in Sweden as observed in separate years, the different years corresponding to different statistical units. The property examined is the claim ratio of the rural fire insurance company in the district considered, the claim ratio being defined as the quotient between the total fire claims paid in the year by the company, and the total sum insured. In this example, the statistical units are modifiable. In fact, we may modify the units in respect of time, and consider, say, the monthly instead of the yearly claim ratios. The modifiability is even double, for we can as well make geographical modifications, i. e. consider the claim ratios in a larger or smaller district.

The modifiability implies that a change in the sizes of the units will give rise to a new statistical population which has a real meaning and is of the same type as the original population. Having stated this, it is evident that we can formally perform similar operations in a population of definite units; the difference is that the modification will in the latter case give rise to fictive units, or units of an essentially different kind than the original ones. In order to illustrate this, we shall consider once more the first example. Let it be assumed that we possess data from a number of Swedish families, and from equally many Norwegian families. We can then exemplify an operation corresponding to the combination of modifiable units by arranging first the families in pairs so

that every pair will contain one Swedish and one Norwegian family, and forming secondly for every such pair the sum of the statures of the two fathers and the sum of the statures of the two sons. Formally, this operation gives rise to a statistical population consisting of the same number of units as each of the two original ones. So far there is a complete analogy with the combination of modifiable units. But the statistical units obtained are altogether fictive, being artificial »families» where e. g. the »stature of the father» is the sum of two real, individual statures.

In an illustration below we shall correlate a series of yearly claim ratios as considered above with a time series which reflects the business cycle. For this purpose, we shall use the yearly number of business failures in the rural districts of Sweden. Evidently, the statistical units will be of the same type as in the second example, i. e. modifiable. The only difference is that we are in this case concerned with another property of the statistical units.

As a final example, let us consider a time series obtained by regular measurements of the air pressure at a certain meteorological station. In this case, each observation refers to a definite time point and to a definite geographical spot. Accordingly, the statistical units are definite. On the other hand, if the observations were given in the form of averages — e. g. the average air pressure during a month, or the average air pressure in a certain district — then the units would obviously be modifiable.

After these examples it is evident that as often as the statistical data refer to time intervals or to geographical areas, the statistical units are modifiable. Moreover, we have seen two type cases of modifiable units. In the first case the statistical data were based on recorded frequencies, in the second the data were averages of recorded measurements.

Thus prepared, we shall next proceed to a more thorough examination of a typical case. It will be found that in a careful analysis of statistical data referring to modifiable units, we must let the hypothetical schemes be influenced by the sizes of the units. In other words, more refined methods will be required than in the classical correlation analysis.

In table 11, cols. (2) — (5) are given three time series of the type examined in Appendix A, i. e. series of claim ratios in Swedish rural fire insurance companies. The series in cols. (2) — (4) have been formed on the basis of all companies which are active in

southern, central, and northern Sweden respectively, while col. (5) refers to the whole of Sweden. With the same geographical classification, cols. (6) — (9) indicate the yearly number of business failures; because of the lag in official registration, these data have been set back one year. Fig. 19 illustrates the data referring to northern Sweden and to the whole of Sweden respectively.

*Table 11. Fire insurance claim ratios (°/oo), and business failures in rural Sweden 1920—34.*

Year	Claim ratios in different districts.				Business failures in different districts.			
	South.	Centr.	North.	Whole	South.	Centr.	North.	Whole
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1920	1.911	.673	.778	.927	379	1 027	1 420	2 826
21	1.871	.953	.761	1.012	413	848	1 612	2 873
22	2.245	.592	.615	.880	381	661	1 086	2 128
23	1.716	.455	.651	.758	320	589	1 072	1 981
24	1.660	.556	.494	.702	260	608	987	1 855
25	1.749	.720	.622	.836	225	627	938	1 790
26	1.734	.648	.764	.878	253	551	881	1 685
27	1.579	.533	.543	.703	316	528	772	1 616
28	1.995	.498	.599	.785	346	536	824	1 706
29	1.586	.673	.622	.758	283	543	753	1 579
1930	1.498	.653	.704	.813	454	551	916	1 921
31	2.733	.648	.689	.999	513	827	1 093	2 433
32	2.702	.572	.708	.972	346	712	984	2 042
33	1.659	.633	.783	.864	266	445	591	1 302
34	1.259	.397	.581	.617	170	292	510	972

According to the previous discussion (see p. 196), we are in table 11 concerned with modifiable statistical units. For instance, the units referred to in col. (5) have been obtained by grouping together the units described in cols. (2) — (4). Of course, owing to the damping effect upon random fluctuations in large statistical populations, we must expect that the combining of districts will give rise to smoother curves. In agreement with this argument, those curves in fig. 19 which refer to northern Sweden are not quite as smooth as those referring to the whole of Sweden. This damping effect of grouping is well-known, and needs no comment. What is of

interest for the moment, and the point of this appendix, is the fact that, considering the correlation between claim ratio and business failures, the grouping has a systematic tendency to increase this correlation. Actually, correlating the claim ratios and the numbers of business failures in the whole of rural Sweden, the resulting correlation coefficient is .750. For southern, central and northern Sweden the coefficients obtained from table 11 are .599, .604 and .375 respectively.

The smoothing effect and the increase of correlation are both

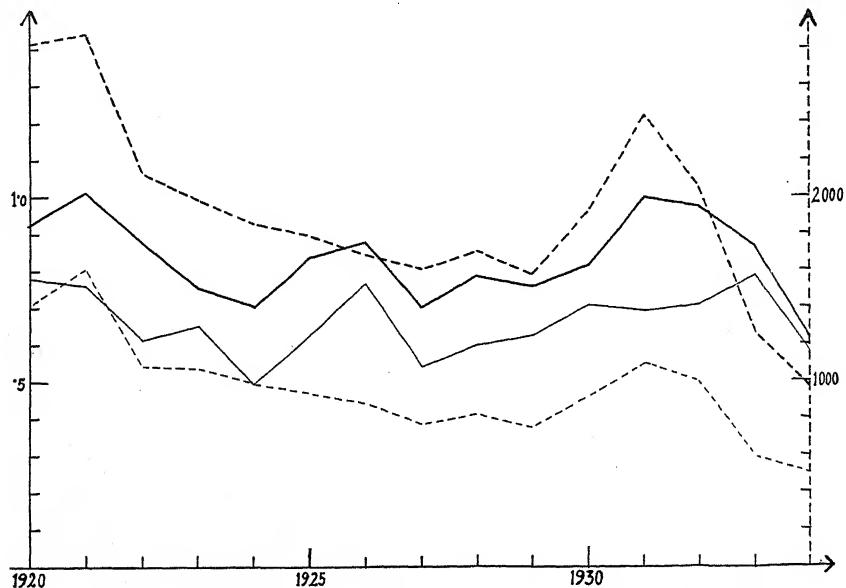


Fig. 19. Fire insurance claim ratios, and business failures in rural Sweden 1920—34. Thick lines: Whole Sweden. Thin lines: Northern Sweden. Unbroken lines: Claim ratios. Broken lines: Business failures.

connected with the law of great numbers. In fact, denoting by  $\xi_t$  and  $\zeta_t$  the claim ratio and the number of business failures, and by  $\{\xi(t)\}$  and  $\{\zeta(t)\}$  two corresponding random processes (see section 11) which are stationary or non-stationary, it is plausible to assume that each of the variables  $\xi(t)$  and  $\zeta(t)$  is composed of a systematic component, say  $\xi^*(t)$  and  $\zeta^*(t)$  respectively, which is unaffected by the grouping, and a random component, say  $\alpha(t)$  and  $\beta(t)$  respectively, which is subject to the law of great numbers. Accordingly, in grouping the data the random components  $\alpha(t)$  and  $\beta(t)$  will decrease relatively to  $\xi^*(t)$  and  $\zeta^*(t)$ . On one hand, in  $\xi(t)$  and

$\zeta(t)$  the trends of  $\xi^*(t)$  and  $\zeta^*(t)$  will then become less blurred by the random components  $\alpha(t)$  and  $\beta(t)$ . This is the smoothing effect. On the other hand, the tendency to an increase in the correlation depends on the structure of the correlation coefficient. Proceeding to prove this, we shall make rather simplifying assumptions.

Retaining the notations, we have

$$(329) \quad \xi(t) = \xi^*(t) + \alpha(t); \quad \zeta(t) = \zeta^*(t) + \beta(t).$$

We shall assume that the processes  $\{\alpha(t)\}$  and  $\{\beta(t)\}$  are purely random and independent of each other and of  $\{\xi^*(t)\}$  and  $\{\zeta^*(t)\}$ . Assuming further  $E[\alpha(t)] = E[\beta(t)] = 0$ , we get  $E[\xi(t) \cdot \zeta(t)] = E[\xi^*(t) \cdot \zeta^*(t)]$  and

$$(330) \quad D^2(\xi) = D^2(\xi^*) + D^2(\alpha), \quad D^2(\zeta) = D^2(\zeta^*) + D^2(\beta).$$

Hence

$$(331) \quad r[\xi(t), \zeta(t)] = \frac{E[\xi(t) \cdot \zeta(t)] - E[\xi] \cdot E[\zeta]}{D(\xi) \cdot D(\zeta)} = \\ = \frac{r[\xi^*(t), \zeta^*(t)]}{\sqrt{1 + \frac{D^2(\alpha)}{D^2(\xi^*)}} \cdot \sqrt{1 + \frac{D^2(\beta)}{D^2(\zeta^*)}}}.$$

Let us now assume that we combine units which are homogeneous in respect to the properties considered. This implies that  $D(\alpha)/D(\xi^*)$  and  $D(\beta)/D(\zeta^*)$  will vary in an inverse proportion to the size of the unit. Thus, these quotients will tend to zero as the unit size tends to infinity. Formula (331) shows that  $r[\xi(t), \zeta(t)]$  in such a case will tend to the limit  $r[\xi^*(t), \zeta^*(t)]$ . On the other hand, if the unit size tends to zero,  $D(\alpha)/D(\xi^*)$  and  $D(\beta)/D(\zeta^*)$  will tend to infinity. The correlation coefficient (331) will then tend to zero.

Of course, the tendencies pointed out above will be present under more general assumptions concerning the components of the variables examined. However, the above discussion is sufficient for our purpose, which is to give a basis for a few general statements concerning correlation analysis of modifiable units.

The situation may be described as follows. *Any correlation index R referring to a modifiable unit is no proper average of the corresponding indices  $R_i$  referring to special parts of the unit. Owing to the relatively larger dispersion of random components at reduced units,  $|R|$  is frequently larger than any  $|R_i|$ .*

Adopting a term proposed by C. SPEARMAN (1907) in a related case (see below) we shall say that the above effect of the division of units is an *attenuation* (*Verdünnung*) of correlation. Conversely, the effect of combining units will be called a *condensation* (*Verdichtung*) of correlation.

We can also describe the situation by saying that *any correlation index referring to a modifiable unit is quantitatively conditioned by the size of the unit referred to*. In other words, *correlation indices referring to units of different sizes are not directly comparable*. For instance, returning to the correlation between fire claims and business failures, the coefficient .375 of northern Sweden is not directly comparable with the figure .750 of whole Sweden, nor with the figures .599 and .604 of southern and central Sweden respectively.

At first glance it might seem paradoxical that a correlation coefficient referring to a geographic district is larger than the coefficients referring to the sub-districts. We have seen that formula (331) resolves the paradox, the explanation being that without further analysis a correlation coefficient referring to a modifiable unit can give only qualitative information as to whether or not a correlation is present. Having found this, our next question will be whether it is at all possible to arrive at quantitative statements about correlation when concerned with modifiable units. In dealing with this problem, we shall attach the analysis to formula (331), and try to estimate the variances of the systematic and the random components. We shall see that in cases such as the above this can be done by means of the variate difference method, and that we in this way can obtain a quantitative interpretation of the correlation between the series examined.

Considering the variables (329), we get without difficulty

$$(332) \quad D^2(\alpha(t)) = D^2(\Delta^k \alpha(t)) / \binom{2k}{k}.$$

Moreover, if the variables  $\xi^*(t)$  are such that, for a certain  $k$ , the differences  $\Delta^k \xi^*(t)$  are approximately vanishing, we get

$$(333) \quad D^2(\Delta^k \xi^*(t)) / \binom{2k}{k} \sim D^2(\alpha).$$

Formulae (332) and (333) form the basis of the variate difference method (see e. g. O. ANDERSON (1929) and G. U. YULE (1921)).

Calculating for  $k = 1, 2, \dots$  the empirical expressions corresponding to (333), viz.

$$(334) \quad \bar{D}^2(\mathcal{A}^k \bar{\xi}_t) / \binom{2k}{k} = \sum_{t=1}^{n-k} (\mathcal{A}^k \bar{\xi}_t)^2 / (n - k) \binom{2k}{k},$$

the values obtained will, roughly speaking, in a few steps decrease to a limit level. This is taken as an estimate of  $D^2(\alpha)$ .

Applying formula (334) to the series given in table 11, and forming the square roots of the resulting values, the first three rows of table 12 were obtained. The columns are numbered as in table 11. In the fourth row are given the dispersions of the original series.

*Table 12. Estimates used in correlation analysis of table 11.*

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$k=1$	.3575	.1017	.0824	.0869	59	97	144	255
$k=2$	.3030	.0801	.0778	.0636	75	62	121	174
$k=3$	.2612	.0601	.0750	.0541	39	54	104	140
$\bar{D}$	.4013	.1246	.0862	.1118	88	171	274	494
$\bar{D}^*$	.3047	.1091	.0425	.0978	65	162	253	474
$q_1$	4.83	1.11	1.39	1	.42	.39	.74	1
$q_2$	3.73	1.46	1.27	1	.42	.57	.71	1

In agreement with the theory of the variate difference method, the first three values in each column — except for the 6th — form a decreasing sequence. Following up the method used, let us fix the values obtained for  $k=3$  as estimates of the dispersions of the random components; in column (6), we choose instead the first value, i.e. 59.

Having fixed the estimates of the random component dispersions, formula (330) supplies us with corresponding estimates say,  $\bar{D}^*$ , of the dispersions of the systematic components. For instance, col. (2) gives  $\bar{D}^* = \sqrt{(4013)^2 - (2612)^2} = .3047$ . The estimates thus obtained are given in the fifth row of table 12.

Observing that formula (331) may be written

$$(335) \quad r^* = r[\xi^*(t), \zeta^*(t)] = \frac{D(\xi) \cdot D(\zeta)}{D(\xi^*) \cdot D(\zeta^*)} \cdot r[\xi(t), \zeta(t)],$$

we get a rough estimate of the left member by inserting the earlier obtained estimates of the quantities appearing in the right member. For instance, the estimates referring to the whole of Sweden give

$$\bar{r}^* \sim \frac{1118}{0.0978} \cdot \frac{494}{474} \cdot (750) = .89.$$

Expressing the result in words, we have performed a hypothetical decomposition of the claim ratio and the number of business failures, so that each series is supposed to consist of a systematic and a random component. The correlation between the two series arises from the correlation between the systematic components. Estimating the correlation coefficient of the systematic components in the two series referring to the whole of Sweden, we have found  $\bar{r}^* \sim .89$ .

It is illuminating to compare the above application of the variate difference method with the well-known application in the trend problem touched upon in the introduction of this appendix. Keeping the notations, the method is in the latter case used for removing the functional components  $x(t)$  and  $y(t)$ , the relevant problem being to study the correlation between the components  $\xi(t)$  and  $\eta(t)$ . In our case, it is  $\xi(t)$  and  $\eta(t)$  that are removed.

We shall next investigate the geographic variation of  $\bar{r}^*$ . Applying formula (335) to the sub-districts referred to in cols. (2)—(4), we get the estimates 1.07, .73 and .82 for southern, central and northern Sweden respectively. In agreement with the above theory, and in contradistinction to the original or *total* correlation coefficients  $\bar{r}$ , the coefficients  $\bar{r}^*$  referring to the whole of Sweden is a genuine average of the coefficients  $\bar{r}^*$  referring to the sub-districts. A short calculation shows that the arithmetical mean of the latter values is .87, a value close to the estimate  $\bar{r}^* = .89$  based on the total material.

The estimates obtained for  $\bar{r}^*$  suggest that there is some geographic variation present in the correlation between the systematic components in the series under analysis. Of course, the estimates of  $\bar{r}^*$  give but a rough idea of this variation, the estimates having been obtained by a very crude method. This is already clear from the fact that for one of the districts the estimate of  $\bar{r}^*$  is above unity.

Finally, we shall touch upon a method to test the estimates yielded by the variate difference method. Considering the random components, it is plausible to assume as a first approximation that

the variance of the random component in the claim sum and in the number of business failures is proportionate to the total claim sum and to the total number of business failures respectively (see e.g. H. CRAMÉR (1927), p. 161). For the claim ratio, we get a corresponding estimate by dividing the total claim sum by the square of the total sum insured. In this way we get an estimate of the relative magnitude of the dispersions of the random components in the different districts, and this estimate is independent of the variate difference method. Writing  $q_2$  for this estimate, the dispersion of the random component in the series referring to the whole of Sweden being taken for unit, I have found for  $q_2$  the values given in the last row in table 12. Writing  $q_1$  for the corresponding values given by the variate difference method, and considering e.g. col. (2), we get  $q_1 = .2612 / .0541 = 4.83$ . The values of  $q_1$  are likewise given in table 12.

The above test of the estimates delivered by the variate difference method suggest that the estimate  $\bar{D}^*$  is too large in col. (2) and somewhat too small in cols. (3) and (7). If we paid regard to this test, we would get a larger dispersion of the systematic component in col. (2), and a smaller one in cols. (3) and (7). The estimate of  $r^*$  yielded by (335) for southern Sweden would then decrease from 1.07 to a value just below unity. In the same way, the estimate obtained for central Sweden would come nearer to the value obtained for the whole of Sweden.

Let us sum up the above analysis. We have started from the hypothesis that the time series examined are composed of a systematic component and a random component. By hypothesis, the former is unaffected when modifying the statistical units, while the latter is subject to a relative damping according to the law of great numbers. This simple and plausible hypothesis is sufficient to explain the decrease in the correlation coefficient when passing to time series referring to smaller geographic districts. In fact, the correlation between the systematic components is the same, but since the division gives rise to relatively larger random components, the total correlation will decrease. On the other hand, the correlation coefficients between the systematic components will be strictly quantitative. Accordingly, the coefficients referring to different geographical districts are directly comparable, whereas the total coefficients are not. Thus, in terms of the hypothetical decomposition we can describe the correlation situation quantitatively.

In cases such as that investigated above, the variate difference method provides us with estimates sufficient for a quantitative correlation analysis. As has been the principle throughout in this volume, the purpose of this application is not to supply a definite theory of the phenomena examined, but merely to illustrate in detail how the methods proposed will work in practice. In the present application, this limitation is the more self-evident, the time series analyzed being very short.

The above analysis covers but a particular case of modifiable units. In other cases, the analysis will require other methods. In the above example, it was possible to apply the variate difference method, a circumstance due to the fact that we were concerned with time series. In spatial series, the modifiability gives rise to more intricate problems. For instance, let us form a spatial series by dividing Sweden into a large number of small sub-districts, and tabulate for each district the claim ratio of a fixed year. Forming a corresponding series of business failures, let us consider the correlation between the two spatial series obtained. The hypothesis of a systematic and a random component being plausible in any series referring to modifiable units, it is evident that even in this case the quantitative significance of the correlation coefficient is conditioned by the sizes of the statistical units. However, the different data in the same series refer to statistical data of different sizes. Consequently, we cannot as before use the variate difference method on the original series to eliminate the systematic component; other methods are required.

Having pointed out the disturbing effect of the random component, it might be asked whether anything general can be said of the magnitude of the effect. Of course, the effect depends on the relative magnitude of the random component. With large statistical masses, the random component often must be relatively small. In such cases the disturbing effect might be neglected, and the correlation coefficients identified with the correlation coefficients between the systematic components.

We shall not carry the analysis further; the purpose of this appendix is not to outline a correlation theory of modifiable units, but rather to state a program, to point out the need of definite hypotheses and of careful and refined methods in the analysis of modifiable statistical units.

In conclusion, a few references will be made to related investigations.

Having searched in the statistical literature for viewpoints related to the above analysis, I have found but one paper to quote, viz. a short note by G. E. GEHLKE and KATHERINE BIEHL (1934). Following a suggestion of H. SHELDON, GEHLKE and BIEHL present certain correlation coefficients which support SHELDON's experience that in spatial series the correlation coefficient has a tendency to increase as districts are combined. — The series examined by GEHLKE and BIEHL refer to modifiable units, and are of the type exemplified on p. 205. The authors give no theoretical analysis of the tendency pointed out. However, our hypothetical decomposition into a systematic and a random component is sufficient to give an explanation of the tendency in question.

Finally, it is of interest to call attention to a formal parallel to our analysis of modifiable units. The hypothetical decomposition into a systematic and a random component is regularly used in the theory of measurements, the accidental errors in the measurements being regarded as purely random. Consequently, studying the correlation between two phenomena, a theoretical analysis leads to a formula of type (331), which shows that the larger the errors, the smaller the observed correlation. P. C. MAHALANOBIS (1922) seems to have been the first to point out this disturbing effect of measurement errors (see also W. PORTIG (1936)). — The C. SPEARMAN (1907) correction for *attenuation* of correlation (cf. also W. BROWN and G. H. THOMSON (1925), Chapter VIII, and T. L. KELLEY (1924), Chapter IX) can be interpreted as a particular case of reckoning with measurement errors.<sup>1</sup> A series of scores in a test of mental abilities of an individual is rather unreliable because of accidental circumstances. This bias can be neutralized by averaging the scores of several tests. In other words, a test series is considered to be composed of a systematic and a random component, the latter being damped when taking the average of different test series. Thus, correlating test series from two individuals, the correlation will be larger when using series obtained by averaging. The ideal case of an infinite number of averaged tests would give the correlation between the systematic components. This correlation can be estimated by applying the SPEARMAN correction of

<sup>1</sup> I am indebted to Professor W. I. THOMAS for this reference.

attenuation, a correction based on a formula of type (331). — Considering a two-dimensional random variable, say  $[\xi, \zeta]$ , and forming a random sample of  $n$  elements, say  $[\xi_1, \zeta_1], \dots, [\xi_n, \zeta_n]$ , S. D. WICKSELL (1918) has studied the problem of estimating the correlation between  $\xi$  and  $\zeta$  on the basis of the correlation in the sample. The formulae of WICKSELL show, i.a., that in general the sampling correlation is smaller, a result which is in agreement with the above analysis, for each sample element may be looked upon as composed of a systematic and a random component.

## List of References.

- O. ANDERSON (1929): *Die Korrelationsrechnung in der Konjunkturforschung. Ein Beitrag zur Analyse von Zeitreihen.* Bonn.
- J. BARTELS (1935): *Zur Morphologie geophysikalischer Zeitfunktionen.* Sitz.-ber. Berliner Wiss., **139**.
- W. H. BEVERIDGE (1921): *Weather and harvest cycles.* Econ. J., **31**.
- KATHERINE BIEHL, see C. E. GEHLKE.
- H. BOHR (1925): *Zur Theorie der fast periodischen Funktionen.* Acta Mathem., **45**.
- W. BROWN and G. H. THOMSON (1921): *The essentials of mental measurement.* 3rd ed. Cambridge.
- B. BRUNO (1938): Article to be published in the *Geografiska Annaler*, **20**.
- H. BRUNS (1911): *Über die Analyse periodischer Vorgänge.* Astron. Nachr., **188**.
- D. BRUNT (1931): *The combination of observations.* 2nd ed. Cambridge.
- (1937): *Climatic cycles.* Geogr. J., **89**.
- H. BURKHARDT (1904): *Trigonometrische Interpolation.* Enzykl. Math. Wiss., Bd II, Teil 1, 1.
- C. H. D. BUYS-BALLOT (1847): *Les changements périodiques de température.* Utrecht.
- F. P. CANTELLI (1916): *La tendenza ad un limite nel senso del calcolo delle probabilità.* R. circ. matem. Palermo, **41**.
- H. S. CARSLAW (1930): *Introduction to the theory of FOURIER's series and integrals.* 3rd ed. London.
- J. I. CRAIG (1916): *A new method of discovering periodicities.* Month. Not. R. Astron. Soc., **76**.
- H. CRAMÉR (1927): *Sannolikhetskalkylen och några av dess användningar.* Stockholm.
- (1928): *On the composition of elementary errors*, I, II. Skandinavisk Aktuarietidskrift, **11**.
- (1937): *Random variables and probability distributions.* Cambridge.
- W. DODGSON, see B. STEWART.
- G. ELFVING (1937): *Zur Theorie der MARKOFFSchen Ketten.* Acta Soc. Sci. Fennicæ (A), **2**.
- R. A. FISHER (1929): *Tests of significance in harmonic analysis.* Proc. Roy. Soc. (A), **125**.
- M. FRÉCHET (1937): *Généralités sur les probabilités. Variables aléatoires.* Paris.

- R. FRISCH (1933): *Propagation problems and impulse problems in dynamic economics.* Economic essays in honour of GUSTAV CASSEL. London.
- C. E. GEHLKE and KATHERINE BIEHL (1934): *Certain effects of grouping upon the size of the correlation coefficient in census tract material.* J. Amer. Stat. Ass., **29**, suppl.
- F. HAUSDORFF (1923): *Momentprobleme für ein endliches Intervall.* Mathem. Zeitschr., **16**.
- G. HERGLOTZ (1911): *Über Potenzreihen mit positivem, reellen Teil im Einheitskreis.* Ber. Leipzig Math.-Phys., **63**.
- T. L. KELLEY (1924): *Statistical method.* New York.
- M. G. KENDALL, see G. U. YULE.
- A. KHINTCHINE (1932): *Sulle successioni stazionarie di eventi.* Giorn. Ist. Ital. Attuari, **3**, 3.
- (1933): *Über stationäre Reihen zufälliger Variablen.* Rec. mathém. Moscou, **40**.
- (1934): *Korrelationstheorie der stationären stochastischen Prozesse.* Mathem. Ann., **109**.
- A. KOLMOGOROFF (1931): *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung.* Math. Ann., **104**.
- (1933): *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Ergebn. Math. u. Grenzgeb., **2**.
- G. KOWALEWSKI (1909): *Einführung in die Determinantentheorie.* Leipzig.
- T. LINDBLAD (1937): *Zur Theorie der Korrelation bei mehrdimensionalen zufälligen Variablen.* Acta Soc. Sci. Fennicæ (A), **2**.
- E. LUNDBERG (1937): *Studies in the theory of economic expansion.* Stockholm.
- P. C. MAHALANOBIS (1922): *On errors of observation and upper air relationships.* Mem. Ind. Meteor. Dep., **24**.
- P. M. MARPLES (1932): *Linear difference equations. An elementary treatment.* J. Inst. Actuaries, **63**.
- R. v. MISES (1931): *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik.* Leipzig und Wien.
- G. MYRDAL (1933): *The cost of living in Sweden 1830—1930.* Stockholm.
- C. W. B. NORMAND (1932): *Present position of seasonal weather forecasting.* Q. J. Roy. Meteor. Soc., **58**.
- S. OPPENHEIM (1909): *Über die Bestimmung der Periode einer periodischen Erscheinung nebst Anwendung auf die Theorie der Erdmagnetismus.* Wien. Sitz.-Ber. (2 a), **118**.
- W. PORTIG (1936): *Messgenauigkeit und Korrelationskoeffizient.* Ann. Hydrogr., **64**.
- G. ROBINSON, see E. T. WHITTAKER.
- V. ROMANOVSKY (1932): *Sur la loi sinusoïdale limite.* R. circ. matem. Palermo, **56**.
- (1933): *Sur une généralisation de la loi sinusoïdale limite.* R. circ. matem. Palermo, **57**.
- C. F. ROOS (1934): *Dynamic economics.* Bloomington, Ind.

- A. SCHUSTER (1898): *On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena.* *Terr. Magn.*, **3**.
- (1900): *The periodogram of the magnetic declination as obtained from the records of the Greenwich Observatory during the years 1871—1895.* *Trans. Cambridge Phil. Soc.*, **18**.
- E. SLUTSKY (1925): *Über stochastische Asymptoten und Grenzwerte.* *Metron*, **5**, 3.
- (1927): *The summation of random causes as the source of cyclic processes* (Russian with an English summary). *Problems of economic conditions*, **3**, 1.
- (1934): *Alcune applicazioni dei coefficienti di FOURIER all'analisi delle funzioni aleatorie stazionarie.* *Giorn. Ist. Ital. Attuari*, **5**, 4.
- (1937): *The summation of random causes as the source of cyclic processes.* *Econometrica*, **5**.
- N. SMIRNOFF (1936): *Sur la distribution de  $\omega^2$ .* *C. R. Acad. Sci. Paris*, **202**.
- C. SPEARMAN (1907): *Demonstration of formulae for the true measurement of correlation.* *Amer. J. Psychol.*, **18**.
- B. STEWART and W. DODGSON (1879): *Preliminary report to the Committee on Solar Physics on a method of detecting the unknown inequalities of a series of observations.* *Proc. Roy. Soc.*, **29**.
- (1879): *Preliminary report to the Committee on Solar Physics on the evidence in favour of the existence of certain short periods common to solar and terrestrial phenomena.* *Proc. Roy. Soc.*, **29**.
- K. STUMPFF (1927): *Analyse periodischer Vorgänge.* Berlin.
- (1936): *Über die Zufalls wahrscheinlichkeit von Periodizitäten in Beobachtungsreihen.* *Veröff. Meteor. Inst. Univ. Berlin*, **1**.
- (1937): *Grundlagen und Methoden der Periodenforschung.* Berlin.
- V. v. SZELISKI (1934): *Analysis of random errors in time series.* Appendix II in C. F. ROOS: *Dynamic economics.* Bloomington, Ind.
- G. H. THOMSON, see W. BROWN.
- J. TINBERGEN (1937): *An econometric approach to business cycle problems.* Paris.
- L. H. C. TIPPETT (1927): *Random sampling numbers.* London.
- SIR G. WALKER (1931): *On periodicity in series of related terms.* *Proc. Roy. Soc. (A)*, **131**.
- E. T. WHITTAKER (1911): *On the law which governs the variations of SS Cygni.* *Month. Not. R. Astron. Soc.*, **71**.
- E. T. WHITTAKER and G. ROBINSON (1926): *The calculus of observations.* 2nd ed. London.
- S. WICKSELL (1918): *On the correlation of acting probabilities.* *Skandinavisk Aktuarie tidskrift*, **1**.
- N. WIENER (1930): *Generalized harmonic analysis.* *Acta Mathem.*, **55**.
- H. WOLD (1935): *A study on the mean difference, concentration curves and concentration ratio.* *Metron*, **12**, 2.
- (1936): *On quantitative statistical analysis.* *Skandinavisk Aktuarie tidskrift*, **19**.

- G. U. YULE (1921): *On the time-correlation problem, with especial reference to the variate-difference correlation method.* J. Roy. Stat. Soc., **84**.
- (1926): *Why do we sometimes get nonsense-correlations between time-series? — A study in sampling and the nature of time-series.* J. Roy. Stat. Soc., **89**.
- (1927): *On a method of investigating periodicities in disturbed series, with special reference to WOLFER's sunspot numbers.* Trans. Roy. Soc. (A), **226**.
- G. U. YULE and M. G. KENDALL (1937): *An introduction to the theory of statistics.* London.

## INDEX.

- ABBE 21.  
Adjustment path 181.  
Aleatory variable 9.  
Alternating errors 195.  
Amplitude 14.  
ANDERSON 201.  
Attenuation of correlation 201.  
Autocorrelation coefficient 12, 33.  
Automoment 33.  
Autoregression, process of 29, 53, 103.  
—»— analysis 80.  
BARTELS 26—27, 73, 151, 189.  
BEVERIDGE 151—152, 154, 158—160,  
162—163, 166—167, 169—171.  
BIEHL 206.  
BOHR 15.  
BOOLE 39—40.  
BOREL 32, 36, 64.  
BROWN 206.  
BRUNO 171—172.  
BRUNS 19, 23, 134.  
BRUNT 119, 171.  
BURKHARDT 13, 24—25.  
BUYS-BALLOT 23—24, 26—27.  
—»— table 23.  
CANTELLI 39.  
CARLEMAN 67.  
CARS LAW 17, 69.  
CESÀRO 73.  
Chance effect 95, 117.  
Characteristic 9, 33.  
—»— equation 20.  
—»— function 42, 62.  
Combination of variables and processes  
10, 32.  
Complete system 187.  
Composed harmonic 15.  
Composition of variables 11.  
Condensation of correlation 201.  
Conditioned variables and characterist-  
ics 34.  
Consistency 32, 70.  
Continuous process 32.  
Convergence in probability 39.  
Convolution 11.  
Correlation coefficient 12.  
Correlogram 135, 147.  
CRAIG 23, 134—139.  
—»— effect 23, 1134, 135—139, 142.  
CRAMÉR 8, 11, 24—25, 36, 43, 61, 66,  
68, 96, 122—123, 148, 152, 191, 204.  
Damped oscillation 20.  
Definite statistical unit 196.  
Discrete process 32.  
Dispersion 9, 10, 33.  
Distributed lag 169.  
Distribution function 9.  
Disturbance 29, 140.  
DODGSON 24.  
Efficiency 6, 72—74, 108, 122, 142, 167,  
178.  
ELFWING 69.  
Equivalence 41.  
EULER 192.  
Evolutive series and processes 1, 96, 140.  
Expectance 22.  
Expectation 10.  
FISHER 119.  
Forecasting 4, 7, 34, 48, 51, 55, 57, 59,  
74, 79—80, 101—103, 106, 122, 135,  
164—165, 168, 171, 186—187.  
FOURIER 1, 6, 17, 58, 66, 73, 118.  
—»— constants 17, 66.  
FRÉCHET 76.  
Frequency 14.  
FRISCH 95, 187.

- Functional approach 14.  
GEHLKE 206.  
Generalized stationary process 34.  
Generating function 67.  
GRAM 76.  
Group of moving averages 127.  
Harmonic analysis 17.  
HAUSDORFF 66.  
HELMERT 21.  
HERGLOTZ 67—68.  
Hidden periodicities process of 22, 58.  
HILBERT 67.  
Homogeneous process 96.  
Independence 11, 37.  
Interpolating process 70.  
KELLEY 206.  
KENDALL 8, 21, 76, 79.  
KHINTCHINE 3—5, 29, 32, 34—35, 58, 63, 66—67, 71.  
KOLMOGOROFF 3—4, 32, 34, 39, 60, 71.  
KOWALEWSKI 45, 76—77, 90.  
Lag 169.  
LARMOR 141.  
Limit process 39.  
LINDBLAD 77.  
Linear approximation 75.  
—»— operations 11, 36, 38.  
—»— regression, process of 27, 52, 99.  
LUNDBERG 188—189.  
MACLAURIN 192.  
MAHALANOBIS 206.  
MARKOFF 69.  
MARPLES 19.  
Mean 9—10, 33.  
MISES 21.  
Model series 49, 52, 56, 59.  
Modifiable statistical unit 8, 30, 109—110, 148, 196.  
Moment 9—10, 67.  
Moving averages, process of 28, 51, 121,  
Multi-dimensional variables and processes  
10, 32.  
MYRDAL, 109, 175—176, 179—180, 183—186, 188.  
Non-autocorrelated 33.  
Non-negative definite 43.  
Normal processes 5, 58, 60—65, 71, 89—92, 131.  
NORMAND 171.  
OPPENHEIM 18—19, 23, 134—139.  
Order of autoregression 54.  
—»— difference equation 96.  
Oscillatory mechanism 94—95.  
PEARSON 21, 193.  
Periodic process 14, 57.  
Periodogram of SCHUSTER 15, 18.  
—»— WHITTAKER 24.  
Periodogram analysis 1, 6, 18, 22—27,  
72—75, 95, 107—108, 114—119, 123,  
134, 142, 146, 148, 151, 166—167,  
171, 177, 188.  
Persistent periodicity 26.  
Phase 14.  
PORTIG 206.  
Primary series and processes 51—52.  
Principal correlation determinant 46.  
—»— distribution function 33.  
Probability function 9.  
Purely random process 21, 47.  
Quadratic form 43, 60.  
Random impulses and shocks 141.  
—»— process 32.  
—»— sample 11.  
—»— sampling numbers 48.  
—»— variable 8—9.  
Rank of singularity 41, 44, 81.  
Regression analysis 76.  
Regular moving average 126.  
Residual variables and processes 77, 82.  
ROBINSON 19, 24, 176.  
ROMANOVSKY 64—65, 120.  
Roos 169.  
Saltus 15, 68.  
Sample series 4, 63.  
—»— value 9.  
SCHMIDT 76.  
SCHUSTER 1, 17, 22—25, 72—73, 107—108, 112.  
SCHWARZ 78, 124—125.  
Serial coefficients 12.  
SHELDON 206.  
Singular variables and processes 41,  
44, 63.  
Sinusoidal limit theorem 2, 28, 64—65, 120—121, 163—166, 174, 185—186.

- SLUTSKY 2, 13, 28—29, 39, 64—65, 73, 109, 118, 120, 187.
- SMIRNOFF 21, 191.
- SPEARMAN 201, 206.
- SPENCER 176.
- Standardized variables 78.
- Stationary process 32.
- Statistical variable 8.
- STEWART 24.
- STIELTJES 9.
- Stochastical difference equation 6—7, 28, 47, 54, 57, 93—99, 104, 187, 190.
- Stochastical variable 9.
- »STUDENT» 25.
- STUMPFF 18, 18, 21—23, 104, 140, 142, 146.
- Superposed harmonics 15.
- », process of 15, 45.
- SZELISKI 195.
- Testing goodness of fit 191.
- »,  $\chi^2$ -method 21.
- »,  $\omega^2$ -method 191.
- Tests of significance 8, 19, 21—23, 25—
- 26, 73, 109—110, 118, 141—143, 148—150, 163—165, 174, 185—186, 203.
- THOMAS 206.
- THOMSON 206.
- TINBERGEN 187.
- TIPPETT 48—49.
- Trend 1, 23, 146, 151, 166, 176, 194.
- Uncorrelated processes 37.
- Variance 9—10.
- Variate difference method 28, 201.
- WALKER 26, 29, 95, 104, 108, 112, 119, 134—135, 140, 143—146, 190.
- WHITTAKER 19, 24—25, 108, 176.
- WICKSELL 206.
- WIENER 16—17, 190.
- WOLD 30, 109, 191.
- WOLFER 29, 140.
- YAMAGA 141.
- YULE 2, 5—6, 8, 12—13, 21, 26, 28—29, 76, 79, 95—96, 105, 108—109, 134—135, 140—143, 148, 166, 173, 178, 185, 189, 201.